TITLE: Target enrichment and extensive population sampling help untangle the recent, rapid radiation of *Oenothera* sect. *Calylophus*

RUNNING HEAD: Cooper et al., Phylogenomics of *Oenothera* sect. *Calylophus*

AUTHORS
Benjamin J. Cooper[1,2,*], Michael J. Moore[3], Norman A. Douglas[4], Warren L. Wagner[5], Matthew G. Johnson[1,6], Rick P. Overson[1,7], Angela J. McDonnell[1], Jeremie B. Fant[1,2], Krissa A. Skogen[1,2], Norman J. Wickett[1,2,*]

[1]The Negaunee Institute for Plant Conservation Science and Action, Chicago Botanic Garden, 1000 Lake Cook Rd., Glencoe, IL 60022
[2]Northwestern University, Program in Plant Biology and Conservation, O.T. Hogan Hall, Room 6-140B, 2205 Tech Drive, Evanston, IL 60208
[3]Oberlin College, Department of Biology, 119 Woodland St., Oberlin, OH 44074
[4] Department of Biology, University of Florida, Gainesville, FL 32611
[5] Department of Botany, MRC-166, Smithsonian Institution, PO Box 37012, Washington, DC 20013-7012
[6] Department of Biological Sciences, Texas Tech University, Box 43131 Lubbock, TX 79409
[7] School of Sustainability, Arizona State University, PO Box 875502, Tempe, AZ 85287-5502
* Authors for correspondence: nwickett@chicagobotanic.org, benjamin.cooper06@gmail.com

ABSTRACT

Oenothera sect. *Calylophus* is a North American group of 13 recognized taxa in the evening primrose family (Onagraceae) with an evolutionary history that may include independent origins of bee pollination, edaphic endemism, and permanent translocation heterozygosity. Like other groups that radiated relatively recently and rapidly, taxon boundaries within *Oenothera* sect. *Calylophus* have remained challenging to circumscribe. In this study, we used target enrichment, flanking non-coding regions, summary coalescent methods, tests for gene flow modified for target-enrichment data, and morphometric analysis to reconstruct phylogenetic hypotheses, evaluate current taxon circumscriptions, and examine character evolution in *Oenothera* sect. *Calylophus*. Because sect. *Calylophus* comprises a clade with a relatively restricted geographic range, we were able to comprehensively sample across the range of geographic and morphological diversity in the group with extensive sampling. We found that the combination of exons and flanking non-coding regions led to improved support for species relationships. We reconstructed potential hybrid origins of some accessions and note that if processes such as hybridization are not taken into account, the number of inferred evolutionary transitions may be artificially inflated. We recovered strong evidence for multiple origins of the evolution of bee pollination from the ancestral hawkmoth pollination, the evolution of edaphic specialization on gypsum, and permanent translocation heterozygosity. This study applies newly emerging techniques alongside dense infraspecific sampling and morphological analyses to effectively address a relatively common but recalcitrant problem in evolutionary biology.

Keywords.– Gypsum Endemism, Onagraceae, *Oenothera* sect. *Calylophus*, Pollinator Shift, Recent Radiation, Phylogenomics, Target Enrichment

INTRODUCTION

The challenges of reconstructing species histories for groups that arose through recent, rapid radiations are well established. Phylogenetic signal can be obscured by processes such as incomplete lineage sorting (ILS) and gene flow (Maddison and Knowles 2006; Knowles and Chan 2008; Christie and Knowles 2015), resulting in short branch lengths and conflicting gene tree topologies. Consequently, approaches that use few loci or concatenation may fail to resolve the most accurate species tree (Eckert and Carstens 2008; Leaché et al. 2014; Xi et al. 2014; Giarla and Esselstyn 2015). This may be particularly common in plants that are thought to have experienced rapid or recent radiation with ongoing hybridization and high levels of ILS , as is likely the case in the evening primrose genus *Oenothera* sect. *Calylophus* (Onagraceae). The application of target enrichment methods that efficiently sequence hundreds of nuclear loci, coalescent-based phylogenetic methods that account for ILS and gene flow, and extensive sampling of morphologically diverse populations across the geographic range may allow for more accurate representations of phylogenetic relationships in this group (Maddison and Knowles 2006; Knowles and Chan 2008; Knowles 2009; Mamanova et al. 2010; Lemmon et al. 2012; Straub et al. 2012; Bryson et al. 2014; Weitemier et al. 2014; Mandel et al. 2014; Stephens et al. 2015; Johnson et al. 2016).

Oenothera sect. *Calylophus* is currently considered to comprise seven species (thirteen taxa) with a center of diversity in western Texas, southern New Mexico, and north-central Mexico (Fig. 1; Towner 1977; Turner and Moore 2014; Wagner, in press). Previous analyses suggest that *Oenothera* sect. *Calylophus* forms a well-supported, morphologically coherent clade with a relatively restricted geographic range (Towner 1977; Levin et al. 2004; Wagner et al. 2007; Turner and Moore 2014; Wagner, in press). However, as with other groups that have experienced rapid radiations, taxon boundaries within *Oenothera* sect. *Calylophus* have been challenging to define, likely due to phenomena such as overlapping morphological boundaries, ongoing introgression, and incomplete lineage sorting.

In the most comprehensive study of the group to date, Towner (1977) circumscribed taxa using morphology, breeding system, geography, and ecology, but it was noted (and our field observations confirm) that characters often overlap among taxa (Towner 1977). Taxa within *Oenothera* sect. *Calylophus* are divided into two easily recognizable subsections: subsect. *Salpingia* and subsect. *Calylophus* (Towner 1977; Wagner et al. 2007). Pollination varies between the two subsections; flowers of subsect. *Salpingia* are adapted to vespertine pollination by hawkmoths, except for *O. tubicula*, which opens in the morning and is primarily pollinated by bees (Towner 1977). Taxa in subsect. *Calylophus* are predominantly bee-pollinated (Towner 1977) and have geographic ranges that partially (or even largely) overlap, resulting in occasional morphologically intermediate populations (Towner 1977). While confounding for morphological-based analyses, this observed pattern of reticulation is consistent with a recent, rapid radiation occurring in parallel with climatic fluctuations and increasing aridity in the region since the Pleistocene (Raven 1964; Towner 1977; Nason et al. 2002; Katinas et al. 2004). Hawkmoth pollination, which is ancestral in the family Onagraceae, and common in *Oenothera* sect. *Calylophus*, is known to result in long-distance pollen movement (Stockhouse 1973; Skogen et al. 2019); therefore, gene flow may have been extensive over the evolutionary history of hawkmoth-pollinated taxa, increasing the chances that processes such as historical introgression may obscure phylogenetic signal in extant plants (Elrich and Raven 1969). With a

phylogenomic approach that samples hundreds of nuclear loci, we may better illuminate both the history of these species and the key evolutionary processes related to speciation in this group.

Understanding speciation in angiosperms remains a fundamental question in evolutionary biology (Barrett et al. 1996; Rajakaruna 2004; van der Niet et al. 2006; Wilson et al. 2007; Crepet and Niklas 2009; Peakall et al. 2010; Xu et al. 2011; Van der Niet and Johnson 2012; Boberg et al. 2014). Section *Calylophus* has an evolutionary history that likely involves changes in reproductive systems (pollinator functional group, breeding system, and autogamy) and edaphic endemism. For example, there are thought to be two independent shifts between pollinators in the section from hawkmoth to bee pollination (Towner 1977; Fig. 1b), despite many studies in other plant groups showing a directional bias in shifts from bee to hummingbird or hawkmoth pollination (Barrett et al. 1996; Wilson et al. 2007; Thomson and Wilson 2008; Tripp and Manos 2008; Barrett 2013). However, pollinator shifts that do not follow this sequence may be more likely when the extent of trait divergence and specialization does not completely inhibit secondary pollinators such as bees, as has been suggested in *Oenothera* sect. *Calylophus* (Stebbins 1970; Tripp and Manos 2008; Van Der Niet et al. 2014). Shifts to autogamy are also frequent across angiosperms and in Onagraceae alone there are an estimated 353 shifts to modal autogamy (Raven 1979). *Oenothera* sect. *Calylophus* also includes at least one autogamous species, *O. serrulata*, which exhibits permanent translocation heterozygosity, a phenomenon in which all chromosomes are translocationally heterozygous (PTH; Towner 1977) (Fig. 1b). While the evolution of PTH has been assessed in molecular phylogenetic analyses across Onagraceae (Johnson et al. 2009), no study to date has examined this transition in a well-sampled clade with extensive population sampling. Lastly, abiotic ecological factors such as edaphic specialization are also known to drive speciation in some groups (Rajakaruna 2004; van der Niet et al. 2006). For example, serpentine endemics represent ~10% of the endemic flora in California even though serpentine soils account for about 1% of terrestrial habitat in the state (Brady et al. 2005). Similarly, the Chihuahuan Desert is comprised of numerous isolated islands of gypsum outcrops and current estimates suggest that at least 235 taxa from 36 different plant families are gypsum endemics (Moore and Jansen 2007; Moore et al. 2014). It is suspected that gypsum endemism has also evolved independently in *Oenothera* sect. *Calylophus* at least twice (Towner 1977; Turner and Moore 2014; Fig. 1b). Ultimately, to understand the role that these transitions have played in shaping the diversity of *Oenothera* sect. *Calylophus*, a robust phylogeny is required.

Here, we use target enrichment, summary coalescent methods, and morphometric analyses to reconstruct a phylogenetic hypothesis, examine previous taxonomic concepts, and resolve the history of pollinator shifts, PTH, and gypsum endemism in *Oenothera* sect. *Calylophus*. Target enrichment is a cost-effective method for sequencing hundreds of loci across hundreds of samples, producing highly informative datasets for phylogenetics (Lemmon et al. 2012; Straub et al. 2012; Mandel et al. 2014; Weitemier et al. 2014; Heyduk et al. 2015; Stephens et al. 2015; Johnson et al. 2016). While target enrichment is generally designed to capture coding regions, a significant proportion of flanking non-coding regions can be recovered (the "splash-zone"; Weitemier et al. 2014). The inclusion of non-coding regions may be particularly informative for recent radiations, since these regions are less constrained by selective pressures and may contain on average more informative sites at shallower time scales (Folk et al. 2015). We included these "splash-zone" regions in our sequence alignments to evaluate their impact on reconstructing lower-order relationships. Importantly, we sampled extensively, including individuals from numerous populations across the geographic and morphological ranges of all thirteen taxa in the section (Fig. 1a). This study presents an example of how

138 combining these molecular techniques with dense sampling and morphological analysis can be
139 used to effectively address a common problem in evolutionary biology.
140
141 METHODS
142
143 A total of 194 individuals spanning the geographic, morphological, and ecological ranges
144 of all 13 recognized taxa in *Oenothera* sect. *Calylophus* [following Towner (1977) and Turner
145 and Moore (2014)] were included in this study (Fig. 1a, S1, S2) along with eight outgroups
146 representing other major sections of *Oenothera* (*Eremia, Gaura, Kneiffia, Lavauxia, Oenothera,*
147 *Pachylophus*, and *Ravenia*) and other genera (*Chylismia* and *Eulobus*) in Onagraceae (S1, S2).
148 DNA was extracted from fresh, silica-dried leaf tissue (S3). staiPTH status was determined for
149 individuals in subsect. *Calylophus* by assessing pollen fertility, when flowers were present, using
150 a modified Alexander stain (Alexander 1969, 1980; S3).
151 Target nuclear loci for enrichment were determined by clustering transcriptome
152 assemblies of *Oenothera serrulata* (1KP accession SJAN) and *Oenothera berlandieri* (1KP
153 accession EQYT). Starting with the 956 phylogenetically informative *Arabidopsis* loci identified
154 by Duarte et al. (2010; S3), we identified 322 homologous, single-copy loci in our clusters and
155 used these in the probe design process. Libraries were enriched for these loci using the MyBaits
156 protocol (Arbor Biosciences, Ann Arbor, MI, USA) and sequenced on an Illumina MiSeq (2 x
157 300 cycles, v3 chemistry; Illumina, Inc., San Diego, California, USA). Raw reads have been
158 deposited at the NCBI Sequence Read Archive (BioProject PRJNA544074; See S3 for details).
159 Reads were trimmed using Trimmomatic (Bolger et al. 2014; S3) and trimmed, quality-filtered
160 reads were assembled using HybPiper (Johnson et al. 2016). From the assembled loci, we
161 produced two datasets: **"exons"** – exon-only alignments, and **"supercontig"** – alignments
162 containing the exon alignment and flanking non-coding regions (the "splash-zone" per
163 Weitemier 2014 and reconstructed using supercontigs produced by HybPiper). We used these
164 two datasets to test the most recent taxonomic circumscription of the group with several
165 methods: (1) phylogenetic inference of concatenated alignments using RAxML (Stamatakis
166 2014; two analyses: exons and supercontigs), (2) ASTRAL-II (Mirarab and Warnow 2015;
167 Sayyari and Mirarab 2016) species tree inference (two analyses: exons and supercontigs), (3)
168 SVD Quartets (Chifman and Kubatko 2014, 2015) species tree inference (one analysis:
169 supercontigs), (4) Phyparts (Smith et al. 2015; one analysis: supercontigs), (5) IQtree (Minh et
170 al. 2018) with both gene and site concordance factors (one analysis: supercontigs).
171 We used HyDe (Blischak et al. 2018) to test for putative hybrid origins of selected taxa
172 and accessions by calculating D-Statistics (Green et al. 2010) for a set of hypotheses (S3). To
173 further characterize population-level processes or genetic structure within sect. *Calylophus*, we
174 extracted and filtered SNPs by mapping individual reads against reference supercontigs (see
175 https://github.com/lindsawi/HybSeq-SNP-Extraction) and used Discriminant Analysis of
176 Principal Components (Jombart et al. 2010) as implemented in the R package *adegenet* (Jombart
177 2008) and the snmf function in the LEA package (Frichot and François 2015) in R (R Core
178 Team, 2020; S3). We evaluated current taxonomic concepts and patterns of morphological
179 variation by measuring character states for morphological structures that have been used
180 historically to discriminate taxa in *Oenothera* sect. *Calylophus* (Towner 1977): Plant height, leaf
181 length (distal), leaf width (distal), leaf length/width ratio (distal), leaf length (basal), leaf width
182 (basal), leaf length/width (basal), sepal length, and sepal tip length (S3). Measurements were
183 made for 125 of the sequenced samples (S11); unfortunately, we were unable to measure all

traits for 73 samples because we did not have access to the herbarium vouchers, or the trait of interest was not captured on the voucher, therefore some samples were dropped from the analysis due to missing values. Finally, the number of transitions and inferred ancestral conditions of reproductive system were mapped onto an ASTRAL species tree, with individuals grouped into species, using the stochastic mapping function in the R package *phangorn* version 2.5.5 (Schliep 2011; S3).

RESULTS AND DISCUSSION

Sequencing and Phylogenetic Results

Sequencing resulted in a total of 80,273,296 pairs of 300-bp reads with an average of 625,323 reads per sample. Following quality filtering, assembly and alignment, we recovered 204 loci that were present in at least 70% of the samples. Across all datasets and analyses, *Oenothera* sect. *Calylophus* was monophyletic. At the subsection level there was strong agreement in topology between concatenation and coalescent-based trees. For example, subsect. *Calylophus* was recovered as sister to subsect. *Salpingia* with strong support across all analyses and *O. toumeyi* [considered by Towner (1977) to be in subsection *Salpingia*; Fig. 2, S4-7] was recovered as sister to subsect. *Calylophus* across all trees, with strong bootstrap support. Within subsect. *Calylophus* there was poor resolution for currently recognized taxa in all analyses, whereas taxon relationships were better resolved in subsect. *Salpingia* (Fig. 2, S4-7). With coalescent-based tree reconstruction, most taxa *sensu* Towner (1977) were recovered as monophyletic with moderate to strong support (Fig. 2, S6-8). In contrast, both the exon and supercontig concatenation trees recovered most currently recognized taxa as non-monophyletic (S4, S5). Given that concatenation has been shown to produce incorrect topologies in the presence of high ILS (Roch and Steel 2015) and that *Oenothera* sect. *Calylophus* underwent recent radiation, we believe the paraphyly of taxa in both concatenation trees might be artifactual. We, therefore, prefer to interpret relationships based on our coalescent-based trees, which comprise the focus for the remainder of the paper (Fig. 2, S6-8).

To understand whether summary coalescent relationships display a consistent signal across the genome, we quantified gene tree and site concordance using Phyparts and IQtree. We found that gene tree concordance was highest at the deepest nodes at the species level where we expected less ILS and more time between speciation events (Fig. 2, S8, S9). Correspondingly, gene tree concordance was lowest at the subspecies level where increased sharing of ancestral alleles and ongoing gene flow are more likely (Fig. 2, S8, S9). For example, within *O. hartwegii* less than 1% of genes were concordant for bifurcations representing all currently recognized taxa at the subspecies level (Fig. 2, S8, S9). For species-level nodes with high support and high gene tree concordance, site concordance was also high; for example, *O. lavandulifolia* had high bootstrap support in summary coalescent trees (BS = 100), high gene tree concordance (Phyparts = 94% concordance, gCF = 74), and high site concordance (sCF = 83; Fig. 2, S9h). For subspecies with high support, but low gene tree concordance, site concordance was moderate. For example *O. hartwegii* subsp. *fendleri* had high bootstrap support (BS = 99), low gene tree concordance (Phyparts = <1% concordance, gCF = 1), and moderate site concordance (sCF = 38; Fig. 2, S9b). For subspecies that were monophyletic in our coalescent-based trees, but that had low support and low gene tree concordance, site concordance was moderate with an average of 35% of sites in agreement for taxa at these nodes (S9[c,f,g,v]). For example *O. hartwegii* subsp. *filifolia* had low bootstrap support, low gene tree concordance (Phyparts = <1% concordance,

gCF = 0), and moderate site concordance (sCF = 36; Fig. 2, S9c). This is an important finding because while individual gene histories can be obscured by ILS, site concordance factors, which may be less constrained and less subject to ILS at shallower evolutionary timescales, provide a key alternative method of support.

In general, topologies of exon and supercontig datasets were similar, with no major differences in clade membership, but the inclusion of the splash-zone increased support at shallow nodes in our trees. However, this trend was not universal. For example, in subsect. *Salpingia*, using the supercontig dataset decreased support slightly for one taxon (*O. hartwegii* subsp. *filifolia*), and in subsect. *Calylophus* it led to paraphyly of another (*O. capillifolia* subsp. *capillifolia*). For six other taxa, our results showed that using supercontigs increased bootstrap support. Therefore, these results demonstrated a net benefit of including flanking non-coding regions for resolving relationships among closely related taxa.

Hybridization and Geneflow

Using concatenated loci from the supercontig dataset, we used HyDe (Blischak et al. 2018) to test for signals of hybridization. We used 552,521 sites and tested 22 hypotheses for either individuals or groups suspected to be of hybrid origin based on field observations of morphological intermediacy, geographic location, and topological position in our coalescent-based trees and found evidence of hybridization in three individuals representing two taxa, both in subsect. *Salpingia*. The highest signal of hybridization, with a gamma value ($\hat{\gamma}$) of .947 suggesting more historic gene-flow, was observed in one individual of *O. tubicula* subsp. *strigulosa* (MJM1916.E). This involved admixture between *O. tubicula* subsp. *tubicula* and the clade consisting of *O. hartwegii* subsp. *hartwegii* and *O. hartwegii* subsp. *maccartii* (Z-score = 5.585, p-value = 0.000, $\hat{\gamma}$ = 0.947; Fig. 3a, S10). We also detected significant levels of hybridization, with $\hat{\gamma}$ ranging from 0.332 to 0.338 suggesting more contemporary gene-flow, in two individuals in *O. hartwegii* subsp. *pubescens*, BJC29 (Z-score = 2.378, p-value = 0.009, $\hat{\gamma}$ = 0.338) and MJM594 (Z-score = 2.094, p-value = 0.018, $\hat{\gamma}$ = 0.332). This more recent gene flow involved admixture between *O. hartwegii* subsp. *pubescens* and the clade consisting of *O. hartwegii* subsp. *hartwegii* and *O. hartwegii* subsp. *maccartii* (Fig. 3a, S10).

The finding that one individual of *O. tubicula* subsp. *strigulosa* may be of hybrid origin is consistent with ongoing gene flow between *O. tubicula* subsp. *strigulosa* and its sister taxon *O. tubicula* subsp. *tubicula*. In our coalescent-based analyses, the two subspecies of *O. tubicula* were not recovered as sister taxa, and this relationship was strongly supported (S9[j-k]). If the two *O. tubicula* taxa arose independently, this would support the hypothesis that bee pollination arose in *Oenothera* sect. *Calylophus* independently three times. However, while the summary coalescent analyses we utilized to estimate phylogenies accounted for ILS in tree estimation, they did not account for ongoing gene flow (Meng and Kubatko 2009; Gerard et al. 2011; Kubatko and Chifman 2019). Our HyDe results may support the hypothesis that *O. tubicula* subsp. *strigulosa* has experienced gene flow from two closely related taxa, and may have hybrid origins resulting from crossing between *O. tubicula* subsp. *tubicula* and *O. hartwegii* subsp. *hartwegii* (Fig. 3a, S10). Therefore, the placement of *O. tubicula* subsp. *strigulosa* as sister to the rest of the *O. hartwegii* species complex in our trees is likely the result of past gene flow and does not represent independent origins of bee pollination in subsect. *Salpingia*. These results undscore the importance of explicitly including tests for hybridization in phylogenetic studies. In the case of these data estimating a species tree given a set of gene trees within a coalescent

framework without considering other, non-ILS sources of signal conflict could artificially inflate the number of inferred evolutionary transitions.

Our HyDe results also suggest that at least some of the morphological intermediacy and overlap among taxa in the group is due to continued, or at least recent, gene flow. For example, both *O. hartwegii* subsp. *pubescens* individuals that are inferred to have significant levels of admixture were collected from morphologically intermediate populations of *O. hartwegii* subsp. *pubescens* and *O. hartwegii* subsp. *hartwegii*. In addition, *O. hartwegii* subsp. *hartwegii* was a parent in all three hybridization events (S10). Thus, gene flow may explain this taxon's non-monophyly in our summary coalescent results. However, despite the often confounding patterns of overlapping morphological variation among closely related taxa in subsect. *Salpingia*, this pattern does not necessarily appear to be the result of admixture, as many of the tests for hybridization based on field observations were not significant (Fig 3a, S10). What is also clear from these results is that much like collecting hundreds of nuclear genes provides a more nuanced picture of phylogenetic signal and taxon relationships, our results show that collecting multiple individuals from across the geographic and morphological ranges is necessary for a more complete picture of relationships among closely related taxa.

After filtering, we extracted a set of 9728 single nucleotide polymorphisms (SNPs) from both coding and non-coding regions. A Discriminant Analysis of Principal Components (DAPC; Fig. 3b) clearly distinguishes subsection *Salpingia* from subsection *Calylophus*, with *O. toumeyi* intermediate between the two, which is consistent with the phylogenetic results presented here. Additionally, the DAPC identifies *O. lavandulifolia* as a distinct genetic cluster from the remaining taxa in subsection *Salpingia*. The overlap between taxa, for example between the remaining taxa in subsection *Salpingia*, is consistent with the high levels of gene tree discordance identified by PhyParts (Fig. 3a). For this latter group of taxa, we computed estimates of ancestry coefficients using sNMF, which suggests a substantial amount of shared ancestral polymorphisms while also showing some evidence of clear genetic structure among taxa (Fig. 3c). Consistent with the phylogenetic analyses, there does not appear to be any clear genetic distinction between *O. hartwegii* subsp. *hartwegii* and *O. hartwegii* subsp. *maccartii*, while *O. hartwegii* subsp. *fendleri*, *O. tubicula* subsp. *tubicula,* and *O. hartwegii* subsp. *filifolia* appear to be largely distinct.

Morphological Analysis

We conducted morphometric Principal Components Analysis (PCA) to determine if morphological patterns were consistent with phylogentic results and to examine if specific characters could be used to diagnose taxa as circumscribed by our phylogenetic analysis. Towner (1977) observed overlapping and confounding patterns of morphological variation among taxa within subsections, as for example in the *O. hartwegii* species complex. Despite this, because some taxa (e.g., *O. hartwegii* subsp. *fendleri*) were strongly supported by our summary coalescent trees we expected that they would be well distinguished in morphometric analysis.

In general, we found that the main traits that separated taxa in subsect. *Salpingia* were leaf traits and plant size, while in subsect. *Calylophus* the main traits that separated taxa were sepal traits. In subsect. *Salpingia*, PC1 accounted for 44.2% of variance in PCA, while PC2 accounted for 28.1% (Fig. 4b). Morphological characters most associated with PC1 were leaf width (distal and basal), plant height, and leaf length/width ratio (distal and basal). Those associated with PC2 were leaf length (distal and basal), sepal length, and sepal tip length (Fig. 4b). In subsect. *Calylophus*, PC1 accounted for 43.8% of explained variance and PC2 accounted for 27.4% (Fig.

4c). The characters most associated with PC1 in subsect. *Calylophus* include leaf length (distal and basal), sepal length, and sepal tip length. Those most associated with PC2 were leaf length/width ratio (distal and basal) and distal leaf width (Fig. 4c).

Our results support Towner's understanding of taxon boundaries by underscoring previous difficulties in identifying individuals in this group based on morphology. We found substantial overlapping morphological variation among currently recognized taxa in both subsections, though some taxa exhibited better grouping than others. The amount of overlap between taxa was not a function of the strength of tree support for a given taxon in our summary coalescent results. For example, *O. hartwegii* subsp. *fendleri*, a taxon that was well supported in our summary coalescent trees, exhibited some of the highest degree of overlap with other taxa in PCA space. Conversely, both *O. hartwegii* subsp. *filifolia* and *O. hartwegii* subsp. *pubescens*, two taxa that formed poorly supported clades in our trees, formed clusters on the outer edges of PCA space and had less overlap than other taxa (Fig. 4b). Interestingly, *O. hartwegii* subsp. *hartwegii*, the taxon that was identified as a parent in all three instances of admixture in our HyDe analysis, also overlaps in PCA with most other taxa in subsect. *Salpingia* (Fig. 4b). This is not surprising given that it is widely distributed in northern Mexico and western Texas, and frequently comes into contact with related taxa resulting in sympatric populations and occasional morphologically intermediate populations.

Implications for Reproductive Systems and Edaphic Endemism

Our results show shifts from hawkmoth to bee pollination likely occurred twice in sect. *Calylophus* (S3) and thus may be more common in Oenothera than previously thought. The strongly supported sister relationship of *O. toumeyi* to remaining subsect. *Calylophus* in our summary coalescent results is consistent with two independent shifts, once in the ancestor of subsect. *Calylophus*, and another in subsect. *Salpingia* on the branch leading to *O. tubicula* (Fig. 2). Independent shifts to bee pollination from hawkmoth pollination are perhaps not surprising considering that within sect. *Calyophus*, hawkmoth-pollinated floral forms exhibit plasticity in hypanthium length and diameter and do not prevent occasional pollination by bees (Lewis et al. in prep; Towner 1977). Hawkmoth-pollinated taxa in sect *Calylophus* exhibit vespertine anthesis, which separates them temporally from diurnal bees, but plasticity in the timing of anthesis is also common among populations (Towner 1977), and hawkmoths are documented to vary spatiotemporally in abundance (Miller 1981; Campbell et al. 1997; Artz et al. 2010). Aditionally, it has been shown that florivore-mediated selection drives floral trait shifts in sect. *Calylophus* towards bee pollinated floral forms (Jogesh et al. 2017; Bruzzese et al. 2019). Plasticity in reproductive traits that allow some continued pollination by bees provides an alternative mode of pollen transfer and may represent a mechanism for ensuring pollination. While studies have shown that premating barriers such as these contribute greatly to reproductive isolation (Stanton et al. 2016), our results show that multiple, indepdendent shifts from hawkmoth to bee pollination and associated morphological changes, such as the shorter corolla length of bee pollinated flowers, may occur in sect. *Calylophus*, and hence may not be a particularly reliable character for diagnosing taxa in this group.

Stochastic mapping (supplemental) suggests that there are multiple origins of permanent translocation heterozygosity (PTH) in sect. *Calylophus* While ring chromosomes are common and found in all taxa in sect. *Calylophus*, PTH is currently known from only one taxon, *O. serrulata*. Because neighboring populations of *O. serrulata* and its putative progenitor O. *capillifolia* subsp. *berlandieri* often resemble each other phenetically, Towner (1977)

hypothesized that *O. serrulata* may have originated multiple times through independent origins of translocation heterozygosity in different geographic regions, and may be best recognized as "a complex assemblage of populations having a common breeding system." However, this has never before been explored in a phylogenetic context, nor has it been clearly demonstrated with phylogenetic studies in Onagraceae. In our summary coalescent trees, all currently recognized taxa in subsect. *Calylophus* were paraphyletic and *O. serrulata* was scattered throughout the subsection (Fig. 2, S6, S7). Although support values are not always high for the positions of various populations of *O. serrulata*, there is at least one well defined, well supported split among populations of *O. serrulata*. In our summary coalescent trees, the two *O. serrulata* accessions from south Texas (MJM 970 & MJM 983) grouped with other south Texas populations of *O. capillifolia* subsp. *berlandieri* with generally strong support (Fig. 2, S6, S7). This relationship was supported in PCA space as well, where MJM 983 was morphologically more similar to the south Texas *O. capillifolia* subsp. *berlandieri* accessions than to other *O. serrulata* (Fig 4C). Our results are therefore consistent with an independent origin of PTH in coastal Texas populations of *O. serrulata*, demonstrating at minimum two origins of PTH (see *Taxonomic Implications* below). However we cannot rule out other independent origins of PTH, for example in the populations of *O. capillifolia* subsp. *berlandieri* occupying sand dunes in western Texas and southeastern New Mexico, or in the other taxa in subsect. *Calylophus*: *O. capillifolia* subsp. *berlandieri*, or the gypsum endemic *O. gayleana*.

Independent origins of gypsum endemism in sect. *Calylophus* are also supported by our analyses (supplemental). Edaphic specialization is a fundamental driver of speciation in plants and contributes greatly to endemism and species diversity in areas with geologically distinct substrates such as gypsum and serpentine outcrops (Kruckeberg 1984; Anacker et al. 2011; Cacho and Strauss 2014; Moore et al. 2014). To date, two gypsum endemic taxa have been described in sect. *Calylophus*, one in each subsection: *O. hartwegii* subsp. *filifolia*, which is relatively widespread on gypsum in New Mexico and trans-Pecos Texas and only rarely grows in sympatry with other taxa, and the recently described *O. gayleana*, which is found in southeastern New Mexico and adjacent western Texas, with disjunct populations in northern Texas and western Oklahoma (Turner and Moore 2014). Despite low support and low gene tree congruence in our analyses, the two gypsum endemic taxa had moderate sCF support (Fig. 2, S9), much like other taxa with similarly low support and high levels of discordance. In addition,while both gypsum endemics overlapped with other taxa in our morphometric analysis, they occupied morphological extremes in PCA space (Fig. 4). Given that other well-supported taxa are also not well differentiated in PCA space, it is not remarkable that the two gypsum endemic taxa were also not more differentiated from other taxa morphologically. Perhaps the strongest evidence in our data for their recognition is that we found no evidence of admixture between either of these gypsum endemic taxa and other closely related taxa (Fig. 3a, S10).

Taxonomic Implications

The most consequential taxonomic result that arises from our analysis is the position of *O. toumeyi*, a member of subsect. *Salpingia*, as sister to subsect. *Calylophus* with strong support, rendering subsect. *Salpingia* paraphyletic (Fig. 2, S9). The current taxonomy groups *O. toumeyi* with *O. hartwegii* due to similar floral and bud characters including large flowers and long floral tubes suggestive of hawkmoth pollination, and rounded buds with long, free sepal-tips (Towner 1977). Because the breeding system is a defining difference in the current circumscription

between the two subsections in sect. *Calylophus*, this result supports abandoning subsections altogether in sect. *Calylophus*.

Within subsect. *Salpingia* our results also suggest the need for revision. While our phylogenetic analyses strongly support the current circumscription of *O. lavandulifolia* (sensu Towner 1977) as a distinct species within subsect. *Salpingia* (Fig. 2, S9), the relationships of the other two species *O. hartwegii* and *O. tubicula* are less clear. Towner (1977) differentiated these two species by the breeding system and grouped the five subspecies of *O. hartwegii* together based on a pattern of reticulate and intergrading variation in which taxa were distinguished from one another by often slight differences in pubescence and leaf shape. Our morphometric analysis confirmed this pattern; however, our phylogenetic results indicated that one taxon, *O. hartwegii* subsp. *fendleri*, shares a closer relationship with the bee pollinated *O. tubicula* subsp. *tubicula* than other taxa in the hawkmoth pollinated *O. hartwegii* species complex (Fig. 2, S6, S7). This relationship was strongly supported and renders *O. hartwegii*, according to the current circumscription, paraphyletic (Towner 1977). Based on strong phylogenetic support for this clade, and its strong morphological distinctiveness as described by Towner (1977), we suggest that *O. hartwegii* subsp. *fendleri* be elevated to the species level along with both races of *O. tubicula* which were equally well supported in phylogenetic analysis and are geographically isolated. Furthermore, our results support a specific distinction for *O. hartwegii* subsp. *filifolia*. While this taxon was poorly supported in our summary coalescent trees (Fig. 2, S6, S7), we found no evidence of hybridization between this taxon and other closely related taxa. In addition, *O. hartwegii* subsp. *fillifolia* is restricted to gypsum. Therefore, we believe that the ecological distinctiveness and lack of gene flow of *O. hartwegii* subsp. *filifolia* with other taxa in the *O. hartwegii* species complex warrants its elevation as a distinct species. In light of these changes, and to maintain consistency in classification in the subsection, we feel that despite the evidence of hybridization of *O. hartwegii* subsp. *pubescens* with *O. hartwegii* subsp. *hartwegii*, it possesses a morphological distinctiveness that is supported by our phylogenetic results. We therefore recommend this taxon be elevated to the species level, while *O. hartwegii* subsp. *hartwegii* and *O. hartwegii* subsp. *maccartii* be retained as is, forming a polytypic species with two subspecies.

In contrast to the relatively clear divisions among taxa in subsect. *Salpingia* in our coalescent trees, none of the four currently recognized taxa in the subsect. *Calylophus* were consistently recovered as monophyletic. For example, *O. capillifolia* subsp. *capillifolia* was monophyletic in our exon-only summary coalescent tree, but not in the "supercontig" tree, and *O. capillifolia* subsp. *berlandieri* and *O. serrulata* were scattered throughout sect. *Calylophus* in both trees, perhaps suggesting widespread gene flow and/or multiple origins of PTH (Fig. 2, S6). Importantly, our results suggest that the circumscription of *O. gayleana* sensu Turner and Moore (2014) should be amended. Specifically, we find that the populations of subsect. *Calylophus* from northern Texas and western Oklahoma that were assigned to *O. gayleana* by Turner and Moore (2014; MJM790-5, BJC71) instead may belong to *O. serrulata* based on both their phylogenetic positions (Fig. 2, S6, S7) and reduced pollen fertility (S12). These north Texas/western Oklahoma populations seem to represent slightly narrower-leaved individuals of *O. serrulata*, which is a common inhabitant of the extensive gypsum outcrops of this area (although it is not restricted to gypsum there).

Finally, our results highlight an unrecognized cryptic taxon within *O. capillifolia* formed by southern Texas coastal populations currently recognized as *O. capillifolia* subsp. *berlandieri*. Towner (1977) described *O. capillifolia* as a polytypic species with two well-differentiated

morphological races. Though he noted the geographic and cytological distinction of the southern Texas coastal populations of *O. capillifolia* subsp. *berlandieri*, these populations were included in *O. capillifolia* subsp. *berlandieri* primarily because of, "completely overlapping morphological variation." In our results, this cryptic clade of southern Texas coastal populations of *O. capillifolia* subsp. *berlandieri* is the most phylogenetically well supported clade in subsect. *Calylophus* and therefore may warrant taxonomic distinction based on our data (Fig. 2, S9r). Similarly, the southern Texas coastal populations of *O. serrulata*, which is likely an independent origin of PTH derived from this cryptic southern Texas coastal clade of *O. capillifolia* subsp. *berlandieri*, are ecologically distinctive and geographically disjunct from other *O. serrulata* (occurring in coastal dunes, unlike other populations in western Texas, Oklahoma, and northern Texas). In the past they were considered distinctive enough to be described as a species, *Calylophus australis* (Towner & Raven 1970). However, Towner (1977) later sunk this variation into *O. serrulata* based on his decision to treat all PTH populations as *O. serrulata*. Combined with our results here and the ecogeographic distinctiveness consistent with an independent origin of PTH in coastal Texas, we believe this taxon also warrants recognition as a second PTH species in *Oenothera* sect. *Calylophus*.

CONCLUSIONS

Here we describe a robust example of resolving a recent, rapid radiation using multiple sources of evidence: (1) extensive sampling from populations throughout the geographic and morphological range, (2) target enrichment for hundreds of nuclear genes, (3) the inclusion of flanking non-coding regions, (4) gene tree-based hybridization inference, (5) SNPs extracted from target enrichment data, and (6) morphometrics. Our results indicate that in recently radiated species complexes with low sequence divergence and/or high levels of ILS that could be an intractable problem with traditional loci, the use of targeted enrichment in addition to flanking non-coding regions provides a net benefit and is essential to recover species-level resolution. Our results also underscore the importance of summary coalescent methods and evaluating gene tree discordance for resolving historical relationships in recalcitrant groups. By explicitly testing for hybridization using gene tree approaches, we also demonstrate that the estimated number of character state transitions may be artifactually inflated if hybridization is not taken into account. This, in combination with morphometrics, provided key evolutionary insights where relationships in summary coalescent methods may be obscured by gene flow. Importantly, our study uncovers strong evidence for multiple origins of biologically important phenomena, including PTH, the evolution of bee pollination, and the evolution of edaphic specialization. Consequently, *Oenothera* sect. *Calylophus* might represent a powerful system for understanding these phenomena, especially with future genome sequence data.

FUNDING SOURCES

This work was supported by National Science Foundation grants DEB-1342873 to KAS, JAF, and NJW, and DEB-1054539 to MJM. Additional support was provided by the National Geographic Society, Oberlin College;, The Chicago botanic Garden, The Shaw Fellowship, New Mexico Native Plant Society, American Society of Plant Taxonomists, and The Society of Herbarium Curators.

REFERENCES

Alexander M.P. 1969. Differential Staining of Aborted and Nonaborted Pollen. Stain Technol. 44:117–122.

Alexander M.P. 1980. A Versatile Stain for Pollen Fungi, Yeast and Bacteria. Stain Technol. 55:13–18.

Anacker B.L., Whittall J.B., Goldberg E.E., Harrison S.P. 2011. Origins and consequences of serpentine endemism in the California flora. Evolution (N. Y). 65:365–376.

Artz D.R., Villagra C. A., Raguso R. A. 2010. Spatiotemporal variation in the reproductive ecology of two parapatric subspecies of *Oenothera cespitosa* (Onagraceae). Am. J. Bot. 97:1498–510.

Barrett S., Harder L., Worley A. 1996. The comparative biology of pollination and mating in floweing plants. Philos. Trans. R. Soc. B Biol. Sci. 351:1271–1280.

Barrett S.C.H. 2013. The evolution of plant reproductive systems: how often are transitions irreversible? Proc R Soc B. 280:20130913.

Blischak P.D., Chifman J., Wolfe A.D., Kubatko L.S. 2018. HyDe: A python package for genome-scale hybridization detection. Syst. Biol. 67:821–829.

Boberg E., Alexandersson R., Jonsson M., Maad J., Ågren J., Nilsson L.A. 2014. Pollinator shifts and the evolution of spur length in the moth-pollinated orchid *Platanthera bifolia*. Ann. Bot. 113:267–275.

Bolger A.M., Lohse M., Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics. 30:2114–2120.

Brady K.U., Kruckeberg A.R., Bradshaw H.D. 2005. Evolutionary ecology of plant adaptation to serpentine soils. Annu. Rev. Ecol. Evol. Syst. 36:243–266.

Bruzzese D.J., Wagner D.L., Harrison T., Jogesh T., Overson R.P., Wickett N.J., Raguso R.A., Skogen K.A. 2019. Phylogeny, host use, and diversification in the moth family Momphidae (Lepidoptera: Gelechioidea). PLoS One. 14:e0207833.

Bryson R.W., Linkem C.W., Dorcas M.E., Lathrop A., Jones J.M., Alvarado-Díaz J., Grünwald C.I., Murphy R.W. 2014. Multilocus species delimitation in the *Crotalus triseriatus* species group (serpentes: Viperidae: Crotalinae), with the description of two new species. Zootaxa. 3826:475–496.

Cacho N.I., Strauss S.Y. 2014. Occupation of bare habitats, an evolutionary precursor to soil specialization in plants. Proc. Natl. Acad. Sci. 111:15132–15137.

Campbell D.R.., Waser N.M.., Melendez-Ackerman E.J.. 1997. Analyzing pollinator-mediated selection in a plant hybrid zone: Hummingbird visitation patterns on three spatial scales. Am. Nat. 149:295–315.

Chifman J., Kubatko L. 2014. Quartet inference from SNP data under the coalescent model. Bioinformatics. 30:3317–3324.

Chifman J., Kubatko L. 2015. Identifiability of the unrooted species tree topology under the coalescent model with time-reversible substitution processes, site-specific rate variation, and invariable sites. J. Theor. Biol. 374:35–47.

Christie M.R., Knowles L.L. 2015. Habitat corridors facilitate genetic resilience irrespective of species dispersal abilities or population sizes. Evol. Appl. 8:454-463.

Crepet W.L., Niklas K.J. 2009. Darwin's second "abominable mystery": Why are there so many angiosperm species? Am. J. Bot. 96:366–81.

Duarte J.M., Wall P.K., Edger P.P., Landherr L.L., Ma H., Pires J.C., Leebens-Mack J.,

dePamphilis C.W. 2010. Identification of shared single copy nuclear genes in *Arabidopsis, Populus, Vitis* and *Oryza* and their phylogenetic utility across various taxonomic levels. BMC Evol Biol. 10:61.

Eckert A.J., Carstens B.C. 2008. Does gene flow destroy phylogenetic signal? The performance of three methods for estimating species phylogenies in the presence of gene flow. Mol. Phylogenet. Evol. 49:832–842.

Ehrlich, P. R., & Raven, P. H. (1969). Differentiation of populations. Science. 165:1228–1232.

Folk R.A., Mandel J.R., Freudenstein J. V. 2015. A protocol for targeted enrichment of intron-containing sequence markers for recent radiations: A phylogenomic example from *Heuchera* (Saxifragaceae). Appl. Plant Sci. 3:1500039.

Frichot E, Mathieu F, Trouillon T, Bouchard G and Fran¸cois O (2014) Fast and efficient estimation of individual ancestry coefficients. Genetics. 196:973-983.

Gerard D., Gibbs H.L., Kubatko L. 2011. Estimating hybridization in the presence of coalescence using phylogenetic intraspecific sampling. BMC Evol. Biol. 11:291.

Giarla T.C., Esselstyn J.A. 2015. The challenges of resolving a rapid, recent radiation: Empirical and simulated phylogenomic of Philippine shrews. Syst. Biol. 64(5):727-740.

Green R., Krause J., Briggs A., Rasilla Vives M., Fortea Pérez F. 2010. A draft sequence of the neandertal genome. Science. 328:710–722.

Heyduk K., Trapnell D.W., Barrett C.F., Leebens-mack J.I.M. 2016. Phylogenomic analyses of species relationships in the genus *Sabal* (Arecaceae) using targeted sequence capture. Biol. J. Linn. Soc. 117(1):106-120.

Jogesh T., Overson R.P., Raguso R.A., Skogen K.A. 2017. Herbivory as an important selective force in the evolution of floral traits and pollinator shifts. AoB Plants. 9(1):plw088.

Johnson M.G., Gardner E.M., Liu Y., Medina R., Goffinet B., Shaw A.J., Zerega N.J.C., Wickett N.J. 2016. HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. Appl. Plant Sci. 4:1600016.

Johnson M.T.J., Smith S.D., Rausher M.D. 2009. Plant sex and the evolution of plant defenses against herbivores. Proc. Natl. Acad. Sci. 106:18079–18084.

Jombart, T. (2008) adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24:1403-1405.

Jombart T, Devillard S and Balloux, F (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genetics. 11:94.

Katinas L., Crisci J., Wagner W., Hoch P. 2004. Geographical diversification of tribes Epilobieae, Gongylocarpeae, and Onagreae (Onagraceae) in North America, based on parsimony analysis of endemicity and track compatibility analysis. Ann. Missouri Bot. 91:159–185.

Knowles L.L. 2009. Estimating species trees: Methods of phylogenetic analysis when there is incongruence across genes. Syst. Biol. 58:463–467.

Knowles L.L., Chan Y.-H. 2008. Resolving species phylogenies of recent evolutionary radiations. Ann. Missouri Bot. Gard. 95:224–231.

Kruckeberg A. 1984. California Serpentines: Flora, Vegetation, Geology, Soils, and Management Problems. Berkeley: Univ of California Press.

Kubatko L.S., Chifman J. 2019. An invariants-based method for efficient identification of hybrid species from large-scale genomic data. BMC Evol. Biol. 19:1–13.

Leaché A.D., Harris R.B., Rannala B., Yang Z. 2014. The influence of gene flow on species tree estimation: a simulation study. Syst. Biol. 63:17–30.

596 Lemmon A.R., Emme S.A., Lemmon E.M. 2012. Anchored hybrid enrichment for massively
597 high-throughput phylogenomics. Syst. Biol. 61:727–744.
598 Levin R., Wagner W., Hoch P. 2004. Paraphyly in Tribe Onagreae : Insights into Phylogenetic
599 Relationships of Onagraceae Based on Nuclear and Chloroplast Sequence Data. Syst. Bot.
600 29:147–164.
601 Maddison W.P., Knowles L. 2006. Inferring Phylogeny Despite Incomplete Lineage Sorting.
602 Evol. Biol. 55:21–30.
603 Mamanova L., Coffey A.J., Scott C.E., Kozarewa I., Turner E.H., Kumar A., Howard E.,
604 Shendure J., Turner D.J. 2010. Target-enrichment strategies for next- generation
605 sequencing. Nat. Methods. 7:111–118.
606 Mandel J.R., Dikow R.B., Funk V. a, Masalia R.R., Staton S.E., Kozik A., Michelmore R.W.,
607 Rieseberg L.H., Burke J.M. 2014. A target enrichment method for gathering phylogenetic
608 information from hundreds of loci: An example from the Compositae. Appl. Plant Sci. 2:1–
609 6.
610 Meng C., Kubatko L.S. 2009. Detecting hybrid speciation in the presence of incomplete lineage
611 sorting using gene tree incongruence: A model. Theor. Popul. Biol. 75:35–45.
612 Miller R.B.. 1981. Hawkmoths and the Geographic Patterns of Floral Variation in *Aquilegia*
613 *caerulea*. Evolution (N. Y). 35:763–774.
614 Minh B.Q., Hahn M.W., Lanfear R. 2020. New methods to calculate concordance factors for
615 phylogenomic datasets. Mol. Biol. Evol. 37(9):2727-2733.
616 Mirarab S., Warnow T. 2015. ASTRAL-II: Coalescent-based species tree estimation with many
617 hundreds of taxa and thousands of genes. Bioinformatics. 31:i44–i52.
618 Moore M., Jansen R. 2007. Origins and biogeography of gypsophily in the Chihuahuan Desert
619 plant group *Tiquilia* Subg. subg. *Eddya*. Syst. Bot. 32:392–414.
620 Moore M.J., Mota J.F., Douglas N.A., Flores Olvera H., Ochoterena H. 2014. Ecology assembly
621 evolution gypsophile floras. In: Rajakaruna N., Boyd R.S., Harris T.B., editors. Plant
622 Ecology and Evolution in Harsh Environments. New York: Nova Science Publishers, Inc. p.
623 97–128.
624 Nason J.D., Hamrick J.L., Fleming T.H. 2002. Historical vicariance and postglacial colonization
625 effects on the evolution of genetic structure in *Lophocereus*, a Sonoran Desert columnar
626 cactus. Evolution 56:2214–2226.
627 Van der Niet T., Johnson S.D. 2012. Phylogenetic evidence for pollinator-driven diversification
628 of angiosperms. Trends Ecol. Evol. 27:353–361.
629 van der Niet T., Johnson S.D., Linder H.P. 2006. Macroevolutionary data suggest a role for
630 reinforcement in pollination system shifts. Evolution 60:1596–1601.
631 Van Der Niet T., Peakall R., Johnson S.D. 2014. Pollinator-driven ecological speciation in
632 plants: New evidence and future perspectives. Ann. Bot. 113:199–211.
633 Peakall R., Ebert D., Poldy J., Barrow R.A., Francke W., Bower C.C., Schiestl F.P. 2010.
634 Pollinator specificity, floral odour chemistry and the phylogeny of Australian sexually
635 deceptive *Chiloglottis* orchids: implications for pollinator-driven speciation. New Phytol.
636 188:437–50.
637 Rajakaruna N. 2004. The Edaphic Factor in the Origin of Plant Species. Int. Geol. Rev. 46:471–
638 478.
639 Raven P.H. 1964. Catastrophic Selection and Edaphic Endemism. Evolution 18:336–338.
640 Raven P.H. 1979. A survey of reproductive biology in Onagraceae. New Zeal. J. Bot. 17:575–
641 593.

R Core Team. 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
Roch S., Steel M. 2015. Likelihood-based tree reconstruction on a concatenation of aligned sequence data sets can be statistically inconsistent. Theor. Popul. Biol. 100:56–62.
Sayyari E., Mirarab S. 2016. Fast coalescent-based computation of local branch support from quartet frequencies. Mol. Biol. Evol. 33(7):1654-68.
Schliep K.P. 2011. phangorn: phylogenetic analysis in R. Bioinformatics. 27(4):592-593.
Skogen K.A., Overson R.P., Hilpman E.T., Fant J.B. 2019. Hawkmoth pollination facilitates long-distance pollen dispersal and reduces isolation across a gradient of land-use change. Ann. Missouri Bot. Gard. 104:495–511.
Smith S.A., Moore M.J., Brown J.W., Yang Y. 2015. Analysis of phylogenomic datasets reveals conflict, concordance, and gene duplications with examples from animals and plants. BMC Evol. Biol. 15:150.
Stamatakis A. 2014. Stamatakis - 2014 - RAxML version 8 a tool for phylogenetic analysis and post-analysis of large phylogenies. 2010–2011.
Stanton K., Valentin C.M., Wijnen M.E., Stutstman S., Palacios J.J., Cooley A.M. 2016. Absence of postmating barriers between a selfing vs. Outcrossing chilean *Mimulus* species pair. Am. J. Bot. 103:1030–1040.
Stebbins G.L. 1970. Adaptive Radiation of Reproductive Characteristics in Angiosperms, I: Pollination Mechanisms. Annu. Rev. Ecol. Syst. 1:307–326.
Stephens J.D., Rogers W.L., Heyduk K., Cruse-Sanders J.M., Determann R.O., Glenn T.C., Malmberg R.L. 2015. Resolving phylogenetic relationships of the recently radiated carnivorous plant genus *Sarracenia* using target enrichment. Mol. Phylogenet. Evol. 85:76–87.
Stockhouse R.E.I. 1973. Biosystematic Studies of *Oenothera* L. Subgenus *Pachylophus*. Univ. Microfilm. Ph.D. thesis. Colorado State University.
Straub S.C.K., Parks M., Weitemier K., Fishbein M., Cronn R.C., Liston A. 2012. Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. Am. J. Bot. 99:349–64.
Swofford, D. L. 2003. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.
Thomson J.D., Wilson P. 2008. Explaining Evolutionary Shifts between Bee and Hummingbird Pollination: Convergence, Divergence, and Directionality. Int. J. Plant Sci. 169:23–38.
Towner H.F., Raven P.H. 1970. A new species and some new combinations in *Calylophus* (Onagraceae). Madrono. 20: 241-245.
Towner H.F. 1977. The Biosystematics of *Calylophus* (Onagraceae). Ann. Missouri Bot. Gard. 64:48–120.
Tripp E.A., Manos P.S. 2008. Is floral specialization an evolutionary dead-end? Pollination system transitions in *Ruellia* (Acanthaceae). Evolution 62:1712–1737.
Turner B.L., Moore M.J. 2014. *Oenothera gayleana* (*Oenothera* sect. *Calylophus*, Onagraceae), a new gypsophile from Texas, New Mexico, and Oklahoma. Phytologia.org. 96:200–206.
Wagner W., Hoch P., Raven P. 2007. Revised Classification of the Onagraceae. Syst. Bot. Monogr. 83:1–240.
Wagner W.L. In press. *Oenothera.* In: Flora of North America Editorial Committee, editor. Flora of North America North of Mexico. New York and Oxford. Vol. 10.
Weitemier K., Straub S.C.K., Cronn R.C., Fishbein M., Schmickl R., McDonnell A., Liston A.

688 2014. Hyb-Seq: Combining Target Enrichment and Genome Skimming for Plant
689 Phylogenomics. Appl. Plant Sci. 2:1400042.
690 Wilson P., Wolfe A.D., Armbruster W.S., Thomson J.D. 2007. Constrained lability in floral
691 evolution: Counting convergent origins of hummingbird pollination in Penstemon and
692 Keckiella. New Phytol. 176:883–890.
693 Xi Z., Liu L., Rest J.S., Davis C.C. 2014. Coalescent versus Concatenation Methods and the
694 Placement of *Amborella* as Sister to Water Lilies. Syst. Biol. 63:919–932.
695 Xu S., Schlüter P.M., Scopece G., Breitkopf H., Gross K., Cozzolino S., Schiestl F.P. 2011.
696 Floral isolation is the main reproductive barrier among closely related sexually deceptive
697 orchids. Evolution. 65:2606–20.

FIGURES

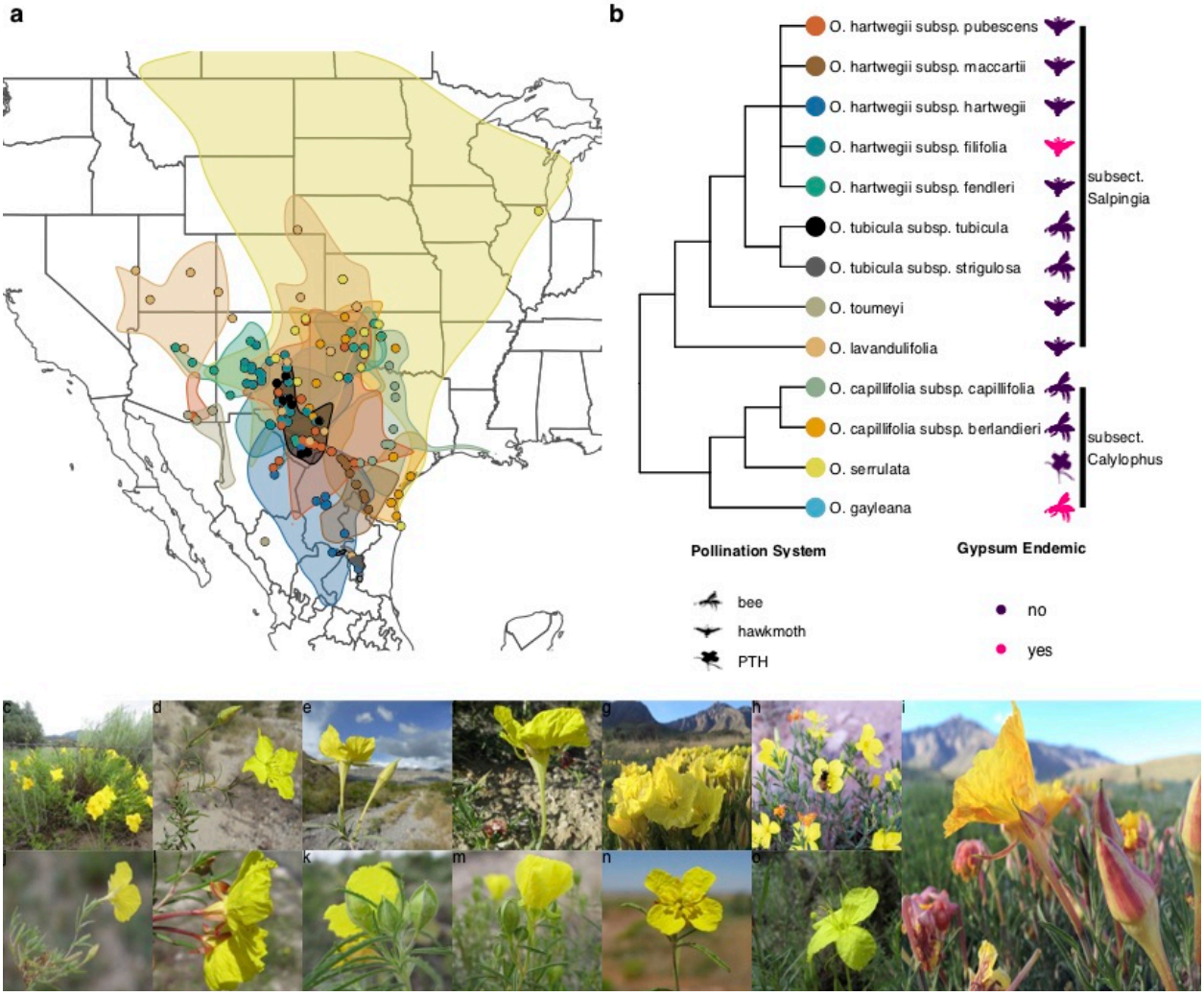


Figure 1. (a) Range map of all taxa in *Oenothera* sect. *Calylophus* (based on Towner 1977). Sampling locations of leaf tissue samples (points; color corresponds to taxa in cladogram to the right [Figure 1b]) and estimated taxon ranges (polygons; colors correspond to Figure 1b) proposed by Towner (1977) and Turner and Moore (2014). (b) Estimated cladogram of *Oenothera* sect. *Calylophus* sensu Towner (1977) and Turner and Moore (2014). Symbols to the right of tip labels signify pollination system (bee, hawkmoth, or Permanent Translocation Heterozygosity [PTH]) and the symbol color specifies whether a given taxon is a gypsum endemic (purple = no, pink = yes). Photo panels: (c) *Oenothera hartwegii* subsp. *fendleri* (d) *Oenothera hartwegii* subsp. *filifolia* (e) *Oenothera hartwegii* subsp. *hartwegii* (f) *Oenothera hartwegii* subsp. *maccartii* (g) *Oenothera hartwegii* subsp. *pubescens* (h) *Oenothera tubicula* subsp. *tubicula* (i) *Oenothera lavandulifolia* (j) *Oenothera tubicula* subsp. *strigulosa* (k) *Oenothera capillifolia* subsp. *berlandieri* (l) *Oenothera capillifolia* subsp. *capillifolia* (m) *Oenothera gayleana* (n) *Oenothera serrulata* (o) *Oenothera toumeyi*

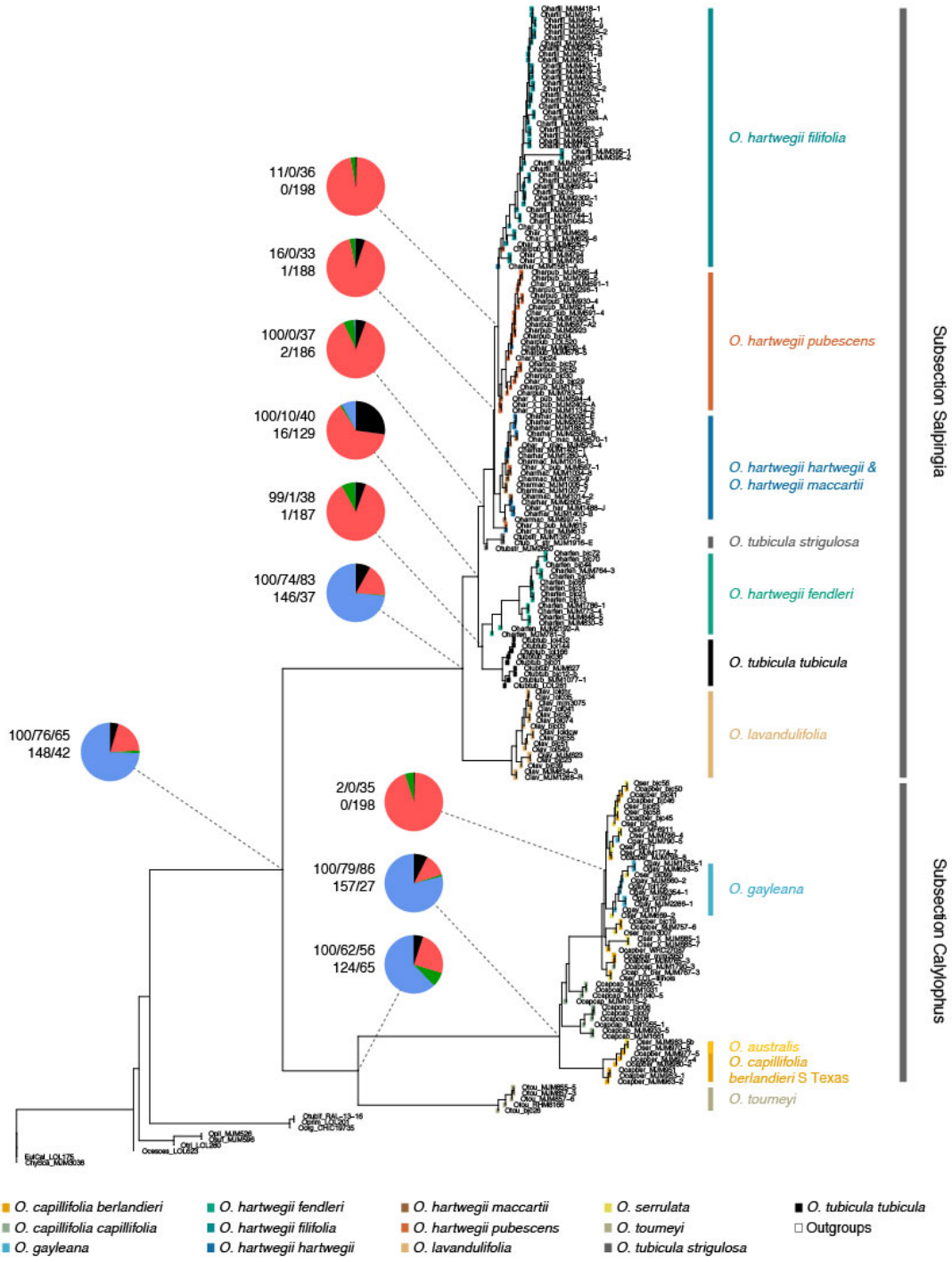


Figure 2. ASTRAL-II summary coalescent tree constructing using the "Supercontig" dataset with 100 bootstraps. At relavent nodes, piecharts represent Phyparts analysis (blue = concordant, green = most conflict, red = all other conflict, black = uninformative gene trees), top row of support values are bootstrap values from ASTRAL-II, and gCF and sCF from IQTree (BS/gCF/sCF), bottom two support values are number of concordant gene trees for the node and total number of gene trees minus the number of concordant gene trees at that node (concord/discord). Colored tip points correspond to taxon designation.

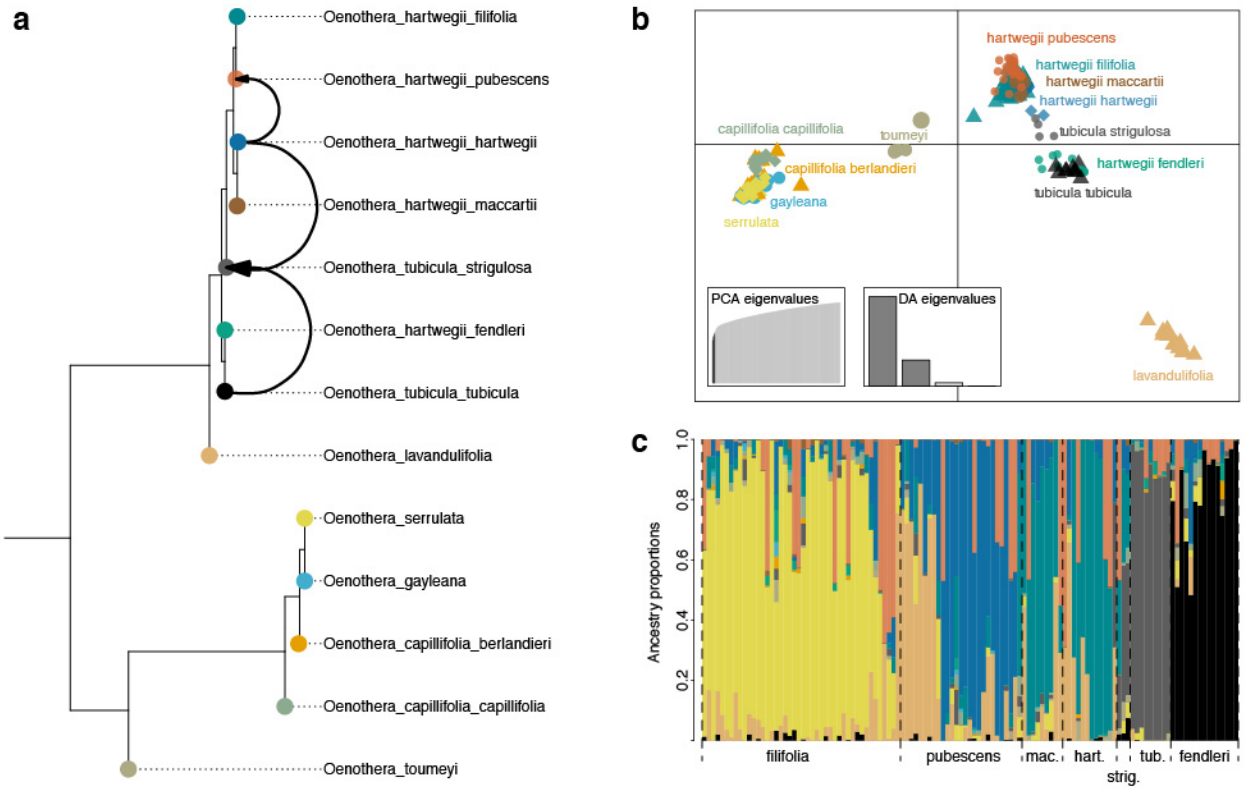


Figure 3. (a) Summary of HyDe Analysis annotated on ASTRAL-III species tree constructed using the "Supercontig" dataset; black arrows represent direction of admixture detected by HyDe anaylsis. (b) Discriminant Analysis of Principal Components based on a filtered set of SNPs extracted from the entire supercontig dataset , and (c) sNMF plot of inferred ancestry coefficients using the same set of filtered SNPs but limited to subsection *Salpingia*.

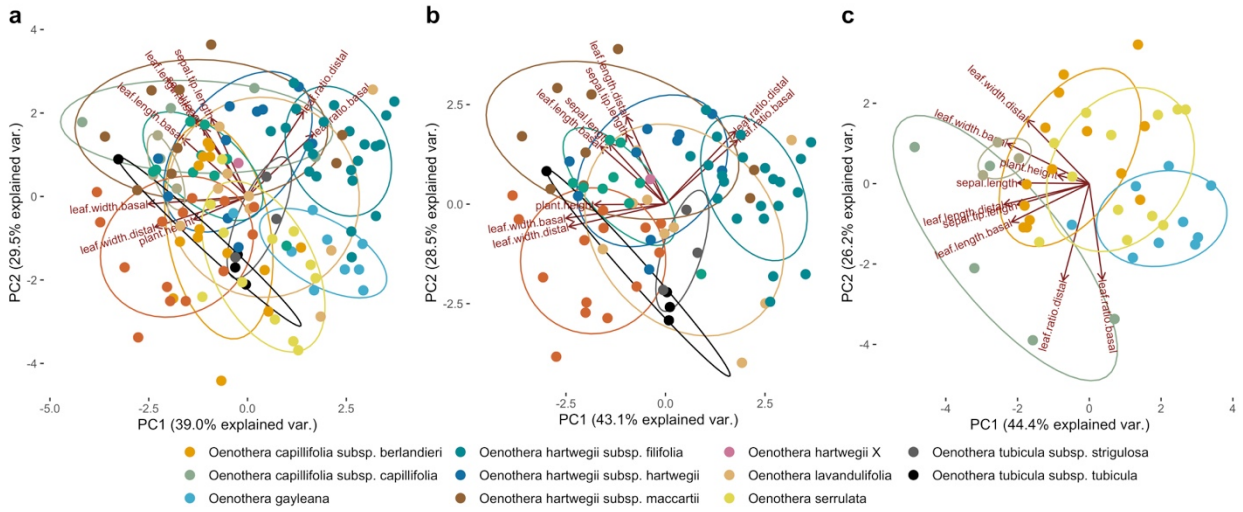


Figure 4. Morphometric Principal Components Analysis (PCA) using 9 morphological characters (plant height, leaf length [basal and distal], leaf width [basal and distal], leaf length/width ratio (basal and distal), and sepal tip length) for (a) Section *Calylophus* (b) Subsection *Salpingia* without *O. toumeyi*, and (c) Subsection *Calylohus* with *O. toumeyi* included.

SUPPLEMENTAL MATERIAL

S1

See excel table "S1 Accessions and Seq Stats"

S2

Number of leaf tissue accessions sequenced from each taxon

Taxon	No. of Accessions
O. lavandulifolia	16
O. tubicula subsp. *strigulosa*	3
O. tubicula subsp. *tubicula*	9
O. hartwegii subsp. *fendleri*	15
O. hartwegii subsp. *filifolia*	44
O. hartwegii subsp. *hartwegii*	12
O. hartwegii subsp. *maccartii*	9
O. hartwegii subsp. *pubescens*	26
O. hartwegii	1
O. toumeyi	5
O. capillifolia subsp. *berlandieri*	17
O. capillifolia subsp. *capillifolia*	11
O. gayleana	9
O. serrulata	17
Outgroups:	
O. pilosella	1
O. organensis	1
O. primiveris	1
O. tubifera	1
O. triloba	1
O. cespitosa subsp. *cespitosa*	1
O. suffrutescens	1
Chylismia scapoidea subsp. *scapoidea*	1
Eulobus californicus	1
Total	203

S3

SUPPLEMENTAL MATERIALS AND METHODS

Taxon Sampling, DNA Extraction, and Determination of PTH

A total of 194 individuals spanning the geographic, morphological, and ecological ranges of all 13 recognized taxa in *Oenothera* sect. *Calylophus* [following Towner (1977) and Turner and Moore (2014)] were included in this study (Fig. 1a, S1, S2) along with eight outgroups representing other major sections of Oenothera (*Eremia*, *Gaura*, *Kneiffia*, *Lavauxia*, *Oenothera*, *Pachylophus*, and *Ravenia*) and other genera (*Chylismia* and *Eulobus*) in Onagraceae (S1, S2). All leaf tissue samples were collected from individuals in the field between 2007 and 2015 and voucher specimens were deposited at the United States National Herbarium (US), with

780 duplicates in most cases at either the Nancy Rich Poole Herbarium (CHIC) or the George T.
781 Jones Herbarium at Oberlin College (OC; S1). DNA was extracted from fresh, silica-dried leaf
782 tissue using either (1) a modified CTAB protocol (Doyle 1987), (2) the Nucleon PhytoPure DNA
783 extraction kit (GE Healthcare Life Sciences, Pittsburgh, Pennsylvania, USA), or (3) a modified
784 CTAB and silicon dioxide purification protocol (Doyle 1987; Sharma and Purohit 2012; See S13
785 for detail) followed by passing any extractions retaining a brown or yellow coloration through a
786 Qiagen Qiaquick PCR spin column for additional purification according to the manufacturer's
787 protocol (Qiagen, Venlo, Netherlands). The third DNA extraction method was used for difficult
788 to extract, polysaccharide-rich leaf tissue samples that yielded gooey, discolored DNA following
789 initial extraction. PTH status was determined for individuals in subsect. *Calylophus* using floral
790 morphology and/or when flowers were present by assessing pollen fertility using a modified
791 Alexander stain, as PTH taxa have a demonstrated 50% reduction in pollen fertility (Towner
792 1977). For accessions identified as either *O. capillifolia* subsp. *berlandieri* or *O. serrulata* that
793 had sufficient pollen available, pollen was removed from flowers and stained using a modified
794 Alexander stain (Alexander 1969, 1980). Accessions with less than 50% viable pollen were
795 assigned to *O. serrulata*, the only currently recognized PTH taxon in subsect. *Calylophus*. Pollen
796 count data are provided in Supplement 12 (See S12 for details).
797
798 *Bait Design, Library Construction, Target Enrichment, and Sequencing*
799 We targeted 322 orthologous, low-copy nuclear loci determined by clustering
800 transcriptomes of *Oenothera serrulata* (1KP accession SJAN) and *Oenothera berlandieri* (1KP
801 accession EQYT) to select a subset of the 956 phylogenetically informative *Arabidopsis* loci
802 identified by Duarte et al. (2010). Transcriptomes of two *Oenothera* species, *O. serrulata* and *O.*
803 *berlandieri*, were assembled and optimal isoforms were filtered for the longest reading frame.
804 Assembled transcripts were aligned as amino acids to the 956 TAIR loci of *Arabidopsis* in
805 TranslatorX (Abascal et al. 2010). This alignment identified 956 orthologous sequences, from
806 which 322 loci were randomly selected. BLAST searches of amino acid sequences from these
807 loci were carried out to ensure orthology between the transcript loci and the *Arabidopsis* TAIR
808 locus. The bait set was designed from these 322 loci, which were selected from both *O. serrulata*
809 and *O. berlanderi* sequences. A set of 19,994 120-bp baits tiled across each locus with a 60 base
810 overlap (2x tiling) was manufactured by Arbor Biosciences (formerly MYcroarray, Ann Arbor,
811 Michigan, USA). Sequencing libraries for 67 samples were prepared with the Illumina TruSeq
812 Nano HT DNA Library Preparation Kit (San Diego, California, USA) following the
813 manufacturer's protocol, except using half volumes beginning with the second addition of
814 Dynabeads MyOne Streptavidin C1 magnetic beads (Invitrogen, Carlsbad, CA, USA). DNA
815 samples were sheared using a Covaris M220 Focused-Ultrasonicator (Covaris, Woburn,
816 Maryland, USA) to a fragment length of ~550 bp (for an average insert size of ~420 bp). The
817 remaining 134 libraries were constructed by Rapid Genomics (Gainesville, Florida, USA), with
818 custom adapters. The Illumina i5 and i7 barcodes were used for all libraries. Target sequences
819 can be accessed at https://github.com/wickettlab/HybSeqFiles.
820 Libraries were enriched for these loci using the MyBaits protocol (ArborBiosciences
821 2016) with combined pools of libraries totaling 1.2 μg of DNA (12 libraries/pool at 100
822 ng/library). Libraries with less than 100 ng of total recovered DNA were pooled together in
823 equimolar concentrations using available product, resulting in some pools with less than 1.2 μg
824 of DNA. The smallest successful pool contained four samples with 6 ng of library each.
825 Hybridization was performed at 65°C for approximately 18 hours. The enriched libraries were

826 reamplified with 14 to 18 PCR cycles and a final cleanup was performed using a Qiagen
827 QiaQuick PCR cleanup kit following the manufacturer's protocol to remove bead contamination
828 (Qiagen, Venlo, Netherlands). DNA concentrations were measured using a Qubit 2.0
829 Fluorometer (Life Technologies, Carlsbad, California, USA) and molarity was measured on an
830 Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, California, USA). A final
831 cleaning step using Dynabeads MyOne Streptavidin C1 magnetic beads was performed on pools
832 with adapter contamination as detected on the Bioanalyzer. Pools were sequenced in four runs at
833 equimolar ratios (4 nM), on an Illumina MiSeq (2 x 300 cycles, v3 chemistry; Illumina, Inc., San
834 Diego, California, USA) at the Pritzker DNA lab (Field Museum, Chicago, IL, USA). This
835 produced 80,273,296 pairs of 300-bp reads. Reads were demultiplexed and adapters trimmed
836 automatically by Illumina Basespace (Illumina 2016). Raw reads have been deposited at the
837 NCBI Sequence Read Archive (BioProject PRJNA544074).
838
839 *Quality Filtering, Assembly and Alignment*
840 A summary of read quality from each sample was produced using FastQC
841 (http://www.bioinformatics.babraham.ac.uk/people.html), which revealed read-through adapter
842 contamination in many of the poorer quality samples. To remove read-through contamination
843 and filter for quality, reads were trimmed for known Illumina adapters using Trimmomatic
844 (Bolger et al. 2014) with the following settings:
845 ILLUMINACLIP:<illumina_adapters.fasta>:2:30:10 LEADING:10 TRAILING:10
846 SLIDINGWINDOW:4:20 MINLEN:20. Trimmed, quality-filtered reads were assembled using
847 the HybPiper pipeline (Johnson et al. 2016) with default settings, followed by the intronerate.py
848 script, to produce both exons and the "splash zone" flanking non-coding region-containing
849 supercontigs. Only pairs with both mates surviving trimming and quality filtering were used for
850 HybPiper.
851 To compare the influence of "splash-zone" non-coding regions, two sets of alignments
852 were created: (1) exons alone, and (2) coding sequences plus the "splash-zone" (Hereby referred
853 to as supercontigs). For multiple sequence alignments of exons alone, protein and nucleotide
854 sequences assembled in HybPiper were gathered into fasta files by gene. For protein sequences
855 only, stop codons were changed to "X" using a sed command-line regular expression to facilitate
856 alignment, and sequences were aligned using MAFFT with settings: --auto --adjustdirection --
857 maxiterate 1000 (Katoh et al. 2002). Aligned protein sequences were then used to fit unaligned
858 nucleotide sequences into coding frame alignments using pal2nal with default settings (Suyama
859 et al. 2006). In-frame, aligned DNA sequences were trimmed to remove low-coverage positions
860 and sequences composed only of gaps using TrimAl with the automated setting, which is
861 optimized for maximum likelihood analyses (Gutíerrez et al. 2009). For supercontigs, nucleotide
862 sequences assembled using HybPiper were gathered into fasta files by gene, gene names were
863 removed from fasta headers using a command-line regular expression, and sequences were
864 aligned in MAFFT with settings: --auto --adjustdirection --maxiterate 1000 (Katoh et al. 2002).
865 Reverse compliment tags ("_R_") were removed from taxon names using a command-line
866 regular expression, and sequences were trimmed using TrimAl with previously listed settings
867 optimized for maximum likelihood analyses (Gutíerrez et al. 2009). To minimize the effects of
868 missing data on phylogenetic analyses, accessions with < 50% of loci passing quality filtering
869 were removed, and genes that were recovered across < 70% of the total remaining samples were
870 also removed. Following quality filtering, we recovered 204 high quality loci (present in at least
871 70% of samples) and an average of 625,323 reads per sample (S1). All pipelines and analyses

872 were run on the high-performance computing cluster at the Chicago Botanic Garden unless
873 otherwise specified.
874
875 *Phylogenetic Reconstruction*
876 We conducted phylogenetic analyses using two strategies for each set of alignments.
877 Alignments were concatenated and analyzed using maximum likelihood (ML) in RAxML
878 (Stamatakis 2014; hereafter referred to as "concatenation"), whereas coalescent-based analyses
879 were conducted using ML gene trees in ASTRAL-II (Mirarab and Warnow 2015; Sayyari and
880 Mirarab 2016) and using unlinked SNPs in SVDquartets (Chifman and Kubatko 2014, 2015)
881 implemented in PAUP* beta version 4.0a168 (Swofford 2003; S3). In concatenation analyses,
882 after aligning each gene separately in MAFFT, genes were concatenated, partitioned, and
883 maximum likelihood trees were reconstructed in RAxML Version 8 (Stamatakis 2014) using the
884 GRTCAT model with 100 "rapid-boostrapping" psuedoreplicates and *Chylismia scapoidea* as
885 the outgroup, on the CIPRES Science Gateway computing cluster (Miller et al. 2010). For
886 coalescent analyses, individual gene trees were first estimated using RAxML Version 8
887 (Stamatakis 2014), with 100 "rapid-bootstraping" psuedo-replicates and settings: -p 12345 -x
888 12345 -N 100 -c 25 -f a -m GTRCAT –s, and *Chylismia scapoidea* as the outgroup. Gene trees
889 based on supercontigs were not partitioned by codon position. Coalescent-based analyses of
890 accessions were conducted in ASTRAL-II (Mirarab and Warnow 2015; Sayyari and Mirarab
891 2016) with default settings using the best RAxML gene trees and their associated bootstrap files
892 as input, and in SVD Quartets (Chifman and Kubatko 2014, 2015) with default settings using an
893 inframe aligned supermatrix with all 204 loci and supercontigs. ASTRAL-II and SVD Quartets
894 analysis was performed with 100 multi-locus bootstraps.
895 To assess concordance among gene trees and provide additional support complementary
896 to bootstrap values, we conducted two additional analyses. First, we assessed raw gene tree
897 concordance using Phyparts (Smith et al. 2015). Prior to running Phyparts, nodes with < 33%
898 support in the supercontig RAxML gene trees were collapsed using the sumtrees command in
899 Dendropy (Sukumaran 2010). These gene trees were then re-rooted using *Chylismia scapoidea*
900 as the outgroup and ASTRAL-II was rerun using these collapsed, re-rooted gene trees as the
901 input files. Pie charts showing gene tree discordance were generated and overlaid on the
902 resulting ASTRAL-II tree using the PhypartsPiecharts script
903 (https://github.com/mossmatters/phyloscripts/tree/master/phypartspiecharts). Phyparts piecharts
904 and gene tree concordance values were also added to Figure 1 by importing the two data files
905 produced by the Phypartspiecharts.py into R and manually matching them to key nodes on our
906 ASTRAL-II supercontig tree using *ggtree* version 1.14.6 (G Yu, DK Smith, H Zhu, Y Guan
907 2017). We also generated gene and site concordance factors for our ASTRAL-II tree constructed
908 using supercontigs in IQTree v1.7-beta16 (Minh et al. 2018). IQtree calculates the gene
909 concordance factor (gCF) and accounts for incomplete taxon coverage among gene trees and
910 therefore may provide a more accurate representation of agreement among gene trees than other
911 methods. In addition to gCF, IQTree calculates the site concordance factor (sCF), which is
912 defined as the percentage of decisive nucleotide sites supporting a specific node (Minh et al.
913 2018). We used the RAxML gene trees produced for the ASTRAL-II supercontig analysis, and
914 supercontig alignments themselves, as the inputs for IQtree. For computing sCF, we randomly
915 sampled 100 quartets around each internal node. Finally, we mapped gCF and sCF values to the
916 ASTRAL-II supercontig tree produced in our previous summary coalescent analysis (Fig. 2). All

phylogenetic trees, with the exception of the full Phyparts picharts tree, were visualized using the R package *ggtree* version 1.14.6 (G Yu, DK Smith, H Zhu, Y Guan 2017).

Ancestral State Reconstruction

To infer ancestral conditions and the number of transitions in reproductive system, we used the *phangorn* (Schliep 2011) package in R. First, a Coalescent-based species tree with accessions grouped into taxa using a mapping file was estimated in ASTRAL-III (Zhang et al. 2018) with default settings using the best RAxML gene trees and their associated bootstrap files, from the supercontig alignments, as input. Next we time calibrated the ASTRAL-III species tree to 1 million years based on estimates of other taxa in the genus (Evans et al. 2009) using the makeChronosCalib function in the *ape* (Paradis et al. 2004) package in R, and estimated an ultrametric tree using the chronos function in *ape* (Paradis et al. 2004) with settings: lambda = 1, model = "relaxed". Finally, we performed marginal reconstruction of ancestral character states using the maximum likelihood method using the optim.pml and ancestral.pml functions in the *phangorn* (Schliep 2011) package in R.

Testing for Hybrid Orgins with HyDe

To test for putative hybrid origins of selected taxa, we used HyDe (Blischak et al. 2018) to calculate D-Statistics (Green et al. 2010) for a set of hypotheses (S10). Briefly, HyDe considers a four-taxon network of an outgroup and a triplet of ingroup populations to detect hybridization from phylogenetic invariants that arise under the coalescent model with hybridization. Introgression between P3 and either P1 or P2 influences the relative frequencies of ABBA and BABA, and the D-statistic measures the imbalance between these frequencies. We tested the triplets in (S10) and set *Chylismia scapoidea* as the outgroup. We considered hypothesis tests significant at an overall $\alpha < 0.05$ level with estimates of γ between 0 and 1. Z-scores greater than 3 are generally interpreted as strong evidence of introgression.

Population-level Analysis

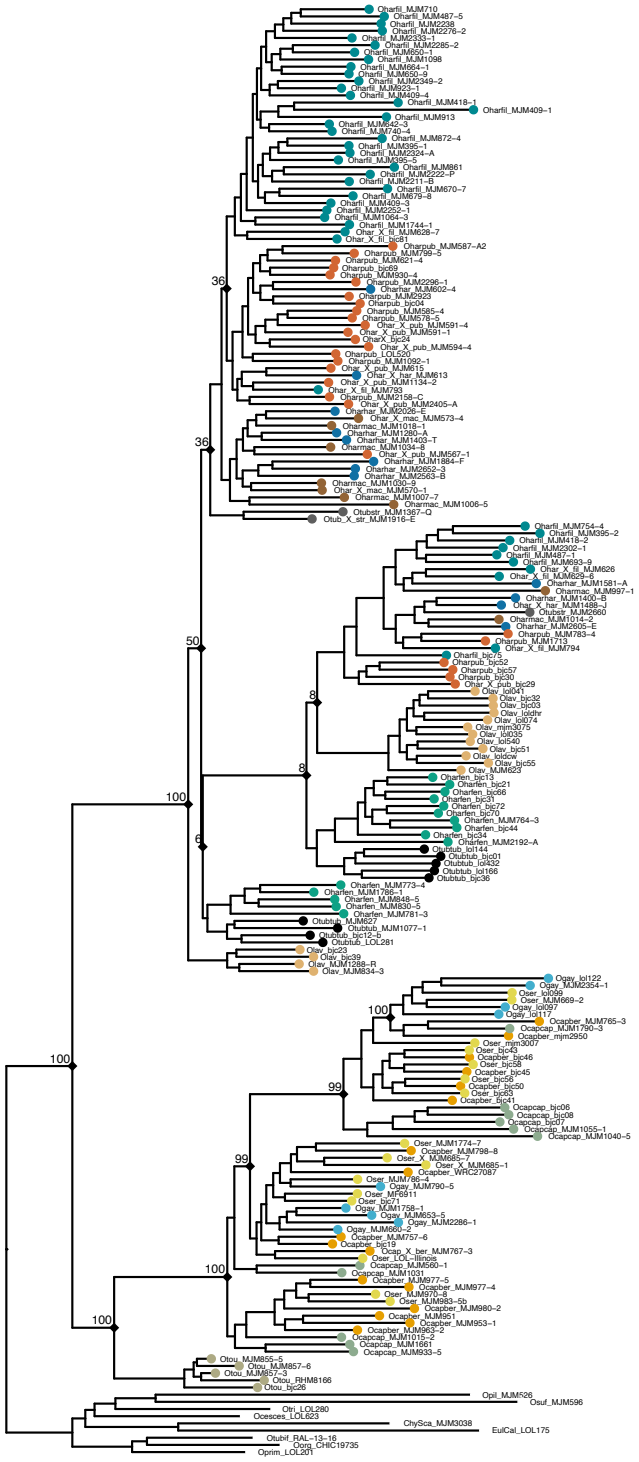
To further characterize population-level processes or genetic structure within sect. *Calylophus*, we extracted and filtered SNPs by mapping individual reads against reference supercontigs (see https://github.com/lindsawi/HybSeq-SNP-Extraction). To account for duplicates arising from PCR during HybSeq in SNP calling and filtering, first we selected the sample with the highest target recovery rate and sequencing depth as a target reference sequence (Oenothera_capillifolia_berlandieri_bjc19) and gathered supercontigs for this individual into a single target FASTA file. We then ran BWA (Li and Durbin 2009) to align sequences, Samtools 'index' (Danecek et al. 2011) and GATK CreateSequenceDictionary (Poplin et al. 2017), respectively, on the resulting target FASTA file. Next we ran a custom script "variant_workflow.sh" using both read files from each *Calylophus* sample as input to create a vcf file for each sample. SNP's were called for each individual using GATK (Poplin et al. 2017) and the vcf file from each sample as input. The resulting vcf file created in the previous step was filtered to remove indels using GATK and the original target FASTA file as input, and then filtered again based on read mapping and quality with GATK VariantFiltration with settings: --filterExpression "QD < 5.0 || FS > 60.0 || MQ < 40.0 || MQRankSum < -12.5 || ReadPosRankSum < -8.0" (Poplin et al. 2017). We generated a reduced SNP file using PLINK (Purcell et al. 2007) to remove SNPs that did pass filter using the command: plink --vcf-filter --vcf input.vcf --const-fid --allow-extra-chr --geno --make-bed --recode structure. Finally we used Discriminant

963 Analysis of Principal Components (Jombart et al. 2010) as implemented in the R package
964 *adegenet* (Jombart 2008) and the snmf function in the LEA package (Frichot and François 2015)
965 in R (R Core Team, 2020) to identify genetic structure between and among individuals and
966 populations in our data.
967
968 *Morphological Measurements and Analysis*
969 To assess taxon boundaries and patterns of morphological variation, we measured
970 character states for the following key morphological structures that have been used historically to
971 discriminate taxa in sect. *Calylophus* (Towner 1977): plant height, leaf length (distal), leaf width
972 (distal), leaf length/width ratio (distal), leaf length (basal), leaf width (basal), leaf length/width
973 (basal), sepal length, and sepal tip length. Measurements were made with digital calipers when
974 possible, or with a standard metric ruler and dissecting scope, from voucher specimens of nearly
975 all sampled populations of sect. *Calylophus* included in our molecular phylogenetic analyses.
976 Measurements are provided in S11. Morphological measurements were log transformed using
977 the R base function 'log' (R Core Team 2018) prior to Principal Components Analysis (PCA),
978 which was conducted in R using the *stats* package version 3.7 and the function 'prcomp'(R Core
979 Team 2018). All 'NA' values were omitted from analysis. Plots of PCA results were visualized
980 using the *ggplot2* package in R (Wickham 2016).
981
982 REFERENCES
983 Abascal F., Zardoya R., Telford M.J. 2010. TranslatorX: Multiple alignment of nucleotide
984 sequences guided by amino acid translations. Nucleic Acids Res. 38:7–13.
985 Alexander M.P. 1969. Differential Staining of Aborted and Nonaborted Pollen. Stain Technol.
986 44:117–122.
987 Alexander M.P. 1980. A Versatile Stain for Pollen Fungi, Yeast and Bacteria. Stain Technol.
988 55:13–18.
989 ArborBiosciences. 2016. In-Solution Sequence Capture for Targeted High-Throughput
990 Sequencing: Version 3.02. .
991 Blischak P.D., Chifman J., Wolfe A.D., Kubatko L.S. 2018. HyDe: A python package for
992 genome-scale hybridization detection. Syst. Biol. 67:821–829.
993 Bolger A.M., Lohse M., Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina
994 sequence data. Bioinformatics. 30:2114–2120.
995 Chifman J., Kubatko L. 2014. Quartet inference from SNP data under the coalescent model.
996 Bioinformatics. 30:3317–3324.
997 Chifman J., Kubatko L. 2015. Identifiability of the unrooted species tree topology under the
998 coalescent model with time-reversible substitution processes, site-specific rate variation,
999 and invariable sites. J. Theor. Biol. 374:35–47.
1000 Danecek P., Auton A., Abecasis G., Albers C.A., Banks E., DePristo M.A., Handsaker R.E.,
1001 Lunter G., Marth G.T., Sherry S.T., McVean G., Durbin R. 2011. The variant call format
1002 and VCFtools. Bioinformatics. 27:2156–2158.
1003 Doyle J.J. 1987. A Rapid DNA Isolation Procedure for Small Quantities of Fresh Leaf Tissue.
1004 Phytochem Bull.:11–15.
1005 Duarte J.M., Wall P.K., Edger P.P., Landherr L.L., Ma H., Pires J.C., Leebens-Mack J.,
1006 dePamphilis C.W. 2010. Identification of shared single copy nuclear genes in *Arabidopsis*,
1007 *Populus, Vitis* and *Oryza* and their phylogenetic utility across various taxonomic levels.
1008 BMC Evol Biol. 10:61.

1009 Evans M.E.K., Smith S. a, Flynn R.S., Donoghue M.J. 2009. Climate, niche evolution, and
1010 diversification of the "bird-cage" evening primroses (*Oenothera*, sections *Anogra* and
1011 *Kleinia*). Am. Nat. 173:225–40.
1012 G Yu, DK Smith, H Zhu, Y Guan T.L. 2017. ggtree: an R package for visualization and
1013 annotation of phylogenetic trees with their covariates and other associated data. Methods
1014 Ecol. Evol.:28–36.
1015 Green R., Krause J., Briggs A., Rasilla Vives M., Fortea Pérez F. 2010. A draft sequence of the
1016 neandertal genome. Science (80-.). 328:710–722.
1017 Gutíerrez S.C., Martínez J.M.S., Gabaldón T. 2009. TrimAl : a tool for automatic alignment
1018 trimming. Bioinformatics.:1–2.
1019 Illumina. 2016. BaseSpace Sequence Hub. Available from https://basespace.illumina.com/.
1020 Johnson M.G., Gardner E.M., Liu Y., Medina R., Goffinet B., Shaw A.J., Zerega N.J.C., Wickett
1021 N.J. 2016. HybPiper: Extracting Coding Sequence and Introns for Phylogenetics from High-
1022 Throughput Sequencing Reads Using Target Enrichment. Appl. Plant Sci. 4:1600016.
1023 Katoh K., Misawa K., Kuma K., Miyata T. 2002. MAFFT: a novel method for rapid multiple
1024 sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30:3059–3066.
1025 Li H., Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform.
1026 Bioinformatics. 25:1754–1760.
1027 Miller M.A., Pfeiffer W., Schwartz T. 2010. Creating the CIPRES Science Gateway for
1028 inference of large phylogenetic trees. 2010 Gatew. Comput. Environ. Work. GCE 2010.
1029 Minh B.Q., Hahn M.W., Lanfear R. 2018. New methods to calculate concordance factors for
1030 phylogenomic datasets. bioRxiv.:487801.
1031 Mirarab S., Warnow T. 2015. ASTRAL-II: Coalescent-based species tree estimation with many
1032 hundreds of taxa and thousands of genes. Bioinformatics. 31:i44–i52.
1033 Paradis E., Claude J., Strimmer K. 2004. APE: Analyses of phylogenetics and evolution in R
1034 language. Bioinformatics. 20:289–290.
1035 Poplin R., Ruano-Rubio V., DePristo M.A., Fennell T.J., Carneiro M.O., Auwera G.A. Van der,
1036 Kling D.E., Gauthier L.D., Levy-Moonshine A., Roazen D., Shakir K., Thibault J.,
1037 Chandran S., Whelan C., Lek M., Gabriel S., Daly M.J., Neale B., MacArthur D.G., Banks
1038 E. 2017. Scaling accurate genetic variant discovery to tens of thousands of samples.
1039 bioRxiv.:201178.
1040 Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A.R., Bender D., Maller J., Sklar
1041 P., De Bakker P.I.W., Daly M.J., Sham P.C. 2007. PLINK: A tool set for whole-genome
1042 association and population-based linkage analyses. Am. J. Hum. Genet. 81:559–575.
1043 R Core Team. 2018. R: A language and environment for statistical computing. .
1044 Sayyari E., Mirarab S. 2016. Fast coalescent-based computation of local branch support from
1045 quartet frequencies. Mol. Biol. Evol.:msw079.
1046 Schliep K.P. 2011. phangorn: Phylogenetic analysis in R. Bioinformatics. 27:592–593.
1047 Sharma P., Purohit S.D. 2012. An improved method of DNA isolation from polysaccharide rich
1048 leaves of Boswellia serrata Roxb. Indian J. Biotechnol. 11:67–71.
1049 Smith S.A., Moore M.J., Brown J.W., Yang Y. 2015. Analysis of phylogenomic datasets reveals
1050 conflict, concordance, and gene duplications with examples from animals and plants. BMC
1051 Evol. Biol. 15:150.
1052 Stamatakis A. 2014. RAxML version 8 a tool for phylogenetic analysis and post-analysis of large
1053 phylogenies. Bioinformatics 30(9):1312-1313.
1054 Sukumaran J. and M.T.H. 2010. DendroPy: A Python library for phylogenetic computing.

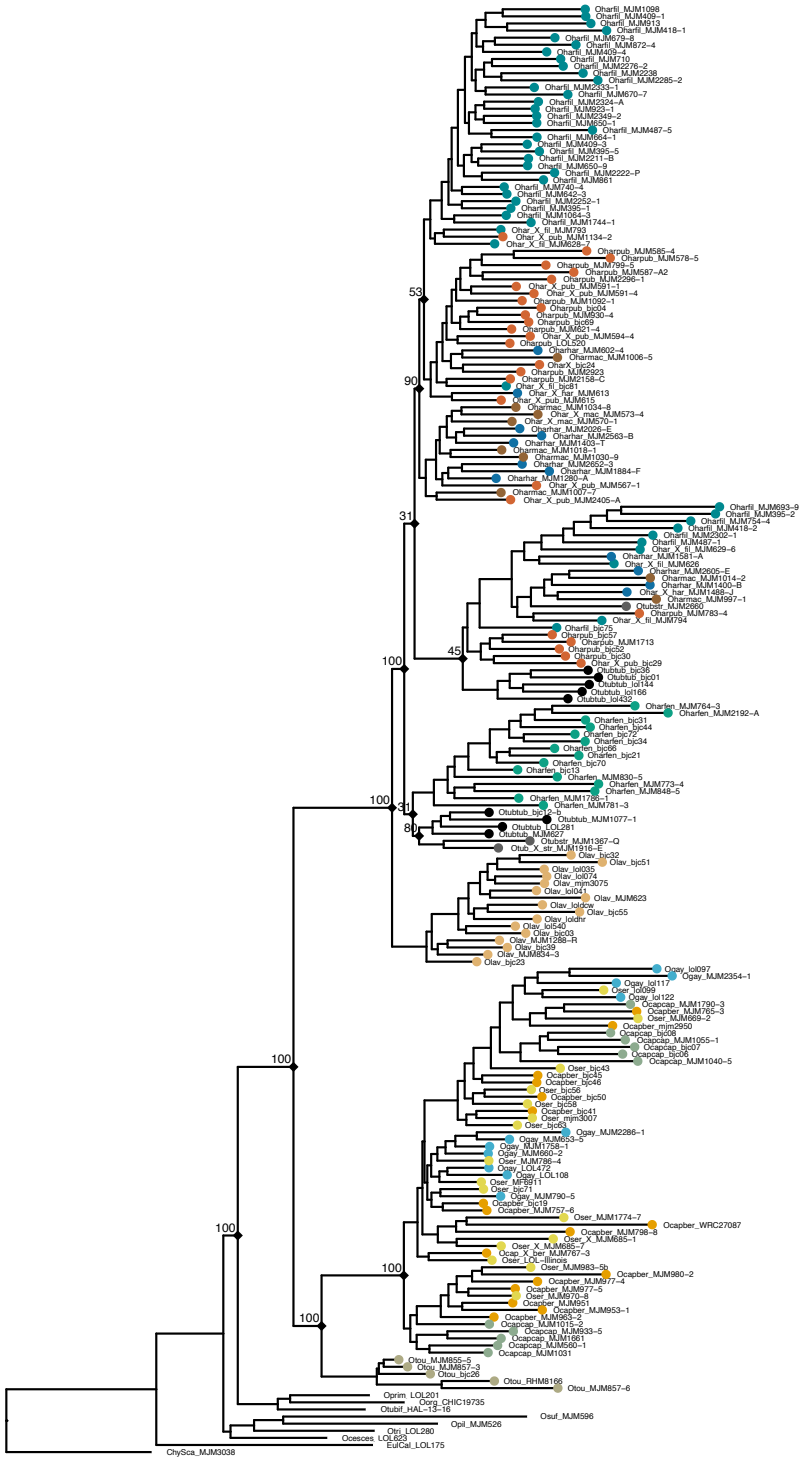
1055 Bioinformatics.:1569–1571.
1056 Suyama M., Torrents D., Bork P. 2006. PAL2NAL: Robust conversion of protein sequence
1057 alignments into the corresponding codon alignments. Nucleic Acids Res. 34:609–612.
1058 Towner H.F.. 1977. The Biosystematics of *Calylophus* (Onagraceae). Ann. Missouri Bot. Gard.
1059 64:48–120.
1060 Wickham H. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag.
1061 Zhang C., Rabiee M., Sayyari E., Mirarab S. 2018. ASTRAL-III: polynomial time species tree
1062 reconstruction from partially resolved gene trees. BMC Bioinformatics. 19:153.

S4



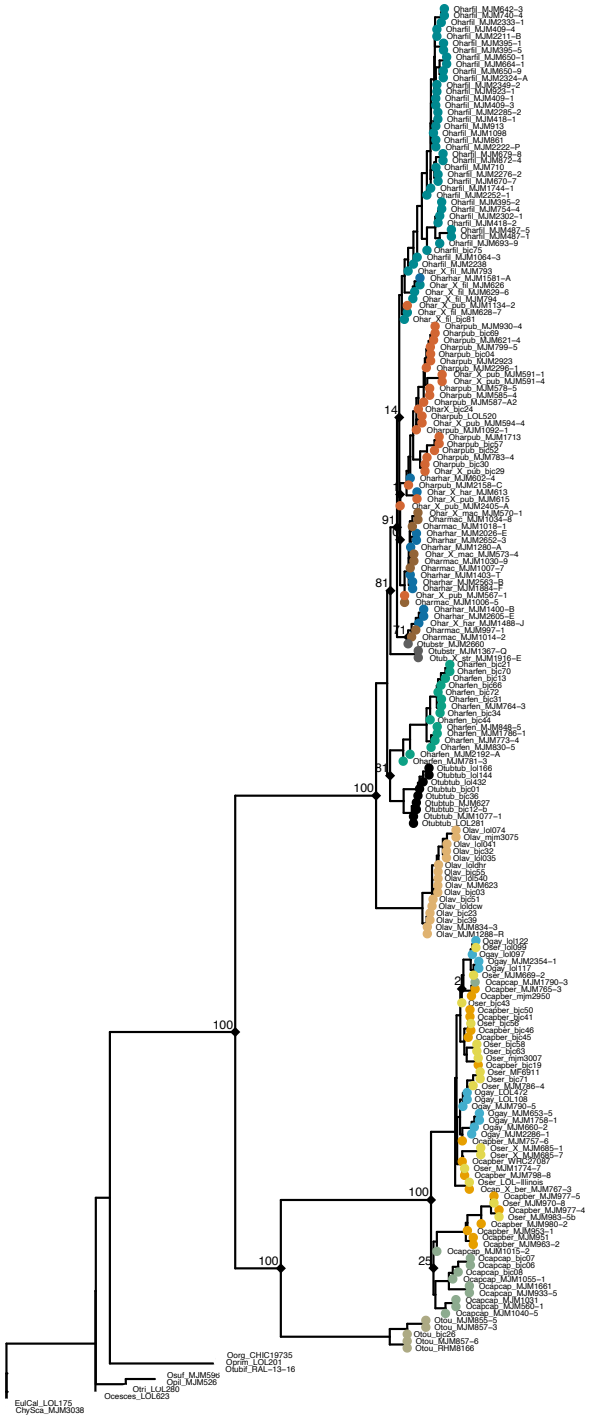
Concatenation tree constructing using the exon-only dataset and 100 bootstraps. Bootstrap values indicated at relevant nodes.

S5



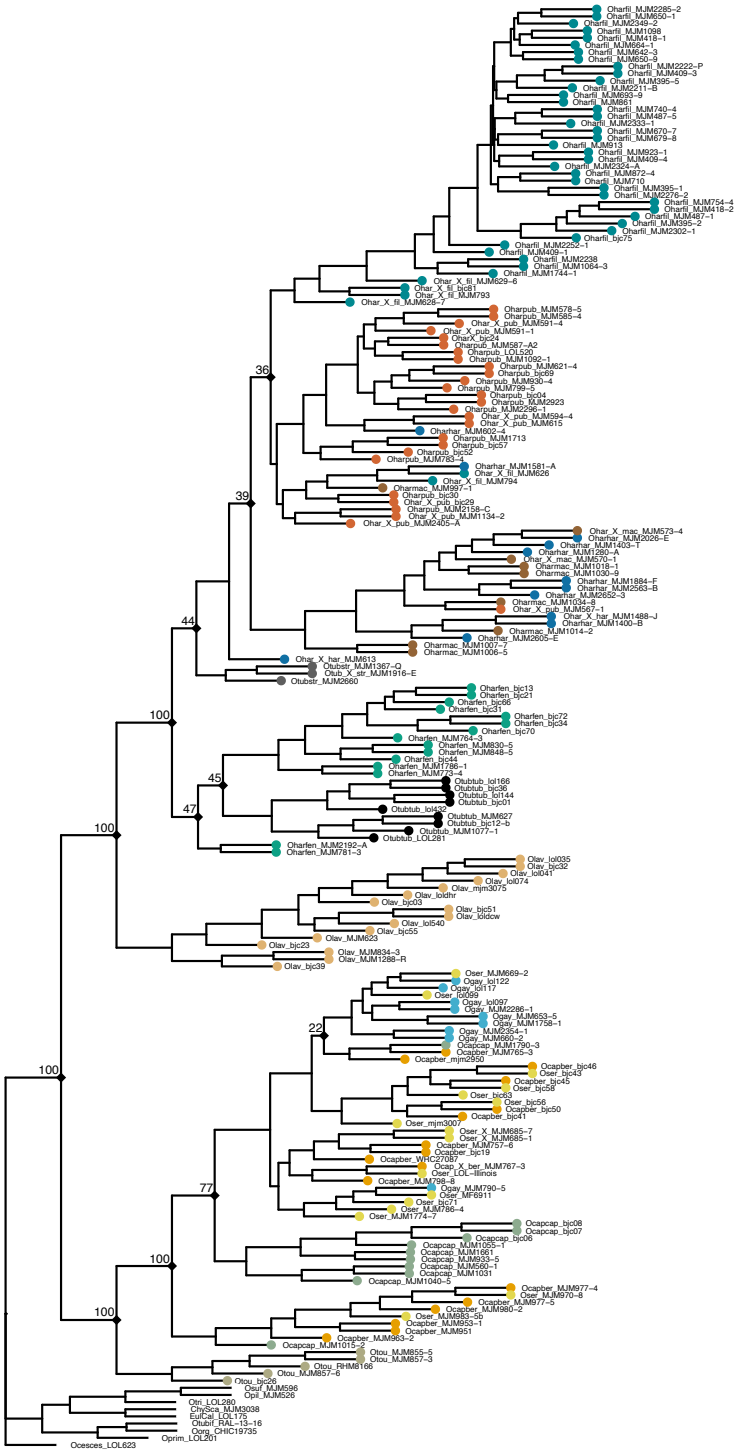
Concatenation tree constructing using the supercontig dataset and 100 bootstraps. Bootstrap values indicated at relevant nodes.

S6



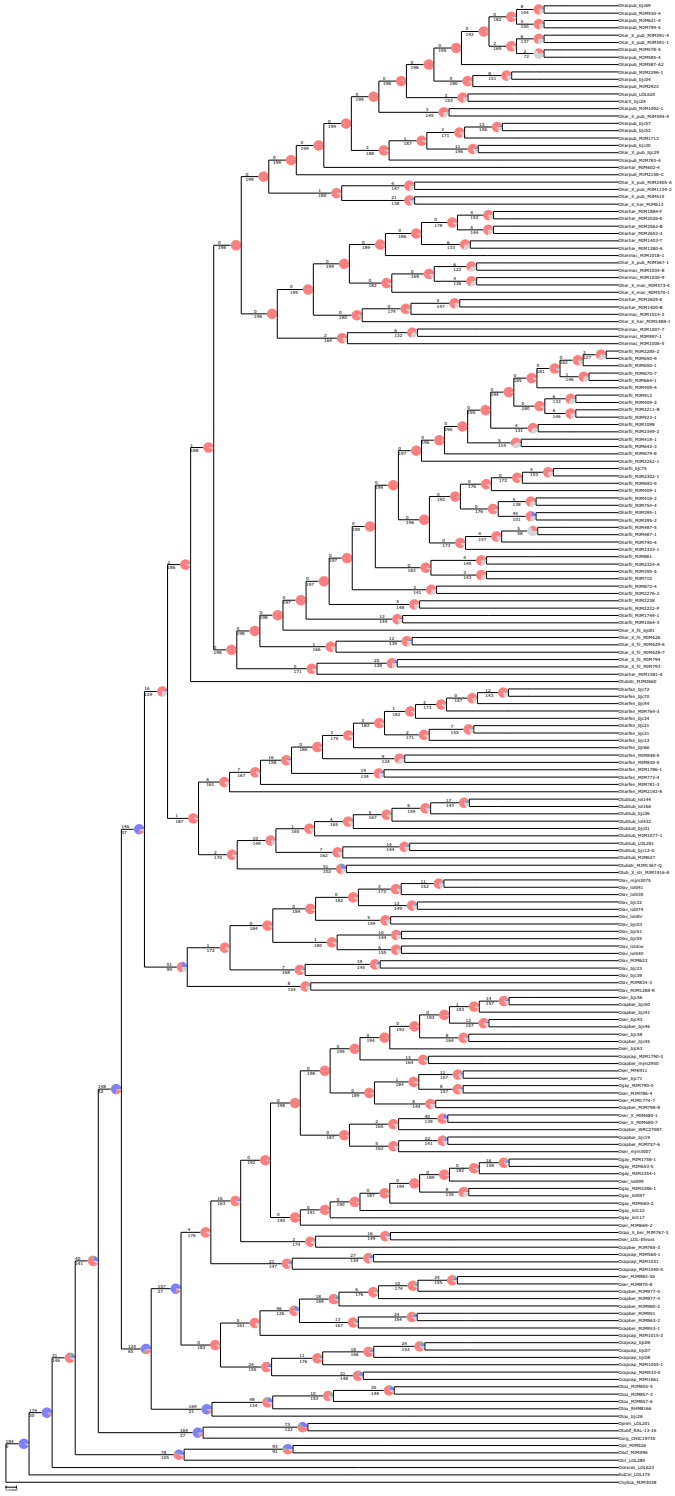
ASTRAL-II summary coalescent tree constructing using the exon-only dataset and 100 bootstraps. Bootstrap values indicated at relevant nodes.

S7



SVD Quartets summary coalescent tree constructing using the supercontig dataset with 100 bootstraps. Bootstrap values indicated at relevant nodes.

S8



Phyparts piecharts ASTRAL-II tree constructing using the supercontig dataset. Piechart colors correspond to: blue = concordant, green = top alternative bipartition, red = all other alternative bipartitions, black = uninformative for that node.

S9

Summary of support values for current and proposed taxonomic treatments, by analysis. ‘e’ signifies trees based on exon-only data, “e+i” signifies trees based on supercontigs. ‘*p’* indicates paraphyletic, unsupported taxon treatment according to tree topology.

	Taxon	Concat. e	Concat. e+i	ASTRAL-II e	ASTRAL-II e+i	SVDQ e+i	Phyparts	gCF iQtree e+i	sCF iQtree e+i
a.	*Oenothera hartwegii* (Towner 1977)	*p*	*p*	*p*	*p*	*p*	*p*	*p*	*p*
b.	*Oenothera hartwegii* subsp. *fendleri* (Towner 1977)	<50	*p*	81	99	*p*	1/ 188	1	38
c.	*Oenothera hartwegii* subsp. *filifolia* (Towner 1977)	*p*	*p*	<50	<50	<50	1/ 190	0	36
d.	*Oenothera hartwegii* subsp. *hartwegii* (Towner 1977)	*p*	*p*	*p*	*p*	*p*	*p*	*p*	*p*
e.	*Oenothera hartwegii* subsp. *macartii* (Towner 1977)	*p*	*p*	*p*	*p*	*p*	*p*	*p*	
f.	*Oenothera hartwegii* subsp. *hartwegii +* subsp. *macartii*	*p*	*p*	<50	<50	<50	0/ 200	0	33
g.	*Oenothera hartwegii* subsp. *pubescens* (Towner 1977)	*p*	*p*	<50	<50	<50	0/ 200	0	36
h.	*Oenothera lavandulifolia* (Towner 1977)	100	*p*	100	100	100	147/37	74	83
i.	*Oenothera tubicula* (Towner 1977)	*p*	*p*	*p*	*p*	*p*	*1/ 188*	*p*	*p*
j.	*Oenothera tubicula* subsp. *tubicula* (Towner 1977)	*p*	*p*	81	99	<50	2/ 171	1	38
k.	*Oenothera tubicula* subsp. *strigulosa* (Towner 1977)	*p*	*p*	81	100	<50	*p*	0	37
l.	*Oenothera fendleri + Oenothera tubicula* subsp. *tubicula*	*p*	*p*	100	100	<50	16/129	10	40
m.	*Oenothera toumeyi* (Towner 1977)	100	100	100	100	100	125/65	62	56
n.	*Oenothera toumeyi + subsect. Calylophus*	100	100	100	100	100	149/42	76	65
o.	*Oenothera capillifolia* (Towner 1977)	*p*	*p*	*p*	*p*	*p*	*p*	*p*	*p*
p.	*Oenothera capillifolia* subsp. *capillifolia* (Towner 1977)	*p*	*p*	<50	*p*	<77	*p*	*p*	*p*
q.	*Oenothera capillifolia* subsp. *berlandieri* (Towner 1977)	*p*	*p*	*p*	*p*	*p*	*p*	*p*	*p*
r.	*Oenothera capillifolia* subsp. *berlandieri* (New Taxon; South Texas coastal populations)	*p*	*p*	<50	100	100	*157/27*	79	86
s.	*Oenothera serrulata* (Towner 1977)	*p*	*p*	*p*	*p*	*p*	*p*	*p*	*p*
t.	*Oenothera ‘australis’* (New Taxon; South Texas *O. serrulata*)	*p*	100	*p*	*p*	*p*	10/175	*p*	*p*
u.	*Oenothera gayleana* (Turner & Moore 2014)	*p*	*p*	*p*	*p*	*p*	*p*	*p*	*p*
v.	*Oenothera gayleana* (Revised Taxon)	*p*	*p*	*p*	<50	<50	0/ 193	0	35

1089 **S10**
1090 List of admixture hypotheses tested, Zscore, P-value and Gamma results from HyDe analysis.
1091 Each row represents a triplet set that was tested consisting of a putative hybrid individual and
1092 two parent groups; Parent 1 and Parent 2.
1093

Putative hybrid individuals	Parent 1 Group	Parent 2 Group	Zscore	P	Gamma
Ohar_X_fil_MJM629.6	CoreOharfil	CoreOharpub	-99999.9	1	-0.049
Ohar_X_fil_MJM628	CoreOharfil	CoreOharpub	-99999.9	1	-0.057
Ohar_X_fil_MJM626	CoreOharfil	CoreOharpub	-99999.9	1	-0.001
Ohar_X_fil_MJM793	CoreOharfil	CoreOharpub	-99999.9	1	-0.057
Ohar_X_fil_MJM794	CoreOharfil	CoreOharpub	-99999.9	1	-0.034
Ohar_X_fil_BJC81	CoreOharfil	CoreOharpub	-99999.9	1	-0.042
Oharhar_MJM1581	CoreOharpub	Oharhar/harmac	-99999.9	1	1.706
Ohar_X_pub_BJC29	CoreOharpub	Oharhar/harmac	2.378	0.009**	0.338
Oharpub_BJC30	CoreOharpub	Oharhar/harmac	-99999.9	1	-0.215
Oharpub_MJM2158.C	CoreOharpub	Oharhar/harmac	1.113	0.133	0.229
Ohar_X_pub_MJM1134.2	CoreOharpub	Oharhar/harmac	0.043	0.483	0.015
Ohar_X_pub_MJM2405.A	CoreOharpub	Oharhar/harmac	0.882	0.189	0.158
Ohar_X_pub_MJM594.4	CoreOharpub	Oharhar/harmac	2.094	0.018*	0.332
Ohar_X_pub_MJM615	CoreOharpub	Oharhar/harmac	0.413	0.340	0.096
Oharmac_MJM997.1	CoreOharpub	Oharhar/harmac	1.563	0.059	0.224
Ohar_X_har_MJM613	CoreOharpub	Oharhar/harmac	-99999.9	1	-0.495
Otub_X_str_MJM1916.E	Otubtub	Oharhar/harmac	5.585	0.000***	0.947
Oharfen_MJM2192.A	CoreOharfen	CoreOharfil	0.305	0.380	0.001
Oharfen_MJM781.3	CoreOharfen	CoreOharpub	0.025	0.490	9.69E-05
Ocapcap_MJM1015.2	CoreOcapcap	S. TX Ocapber	0.467	0.320	0.003
Ocap_X_ber_MJM767.3	central OK Ocapber	Pecos River/Southern Plains	-99999.9	1	-0.003

- Indicates value meets significance threshold of $P < 0.05$
** Indicates value meets significance threshold of $P < 0.001$
*** Indicates value meets significance threshold of $P < 0.0001$

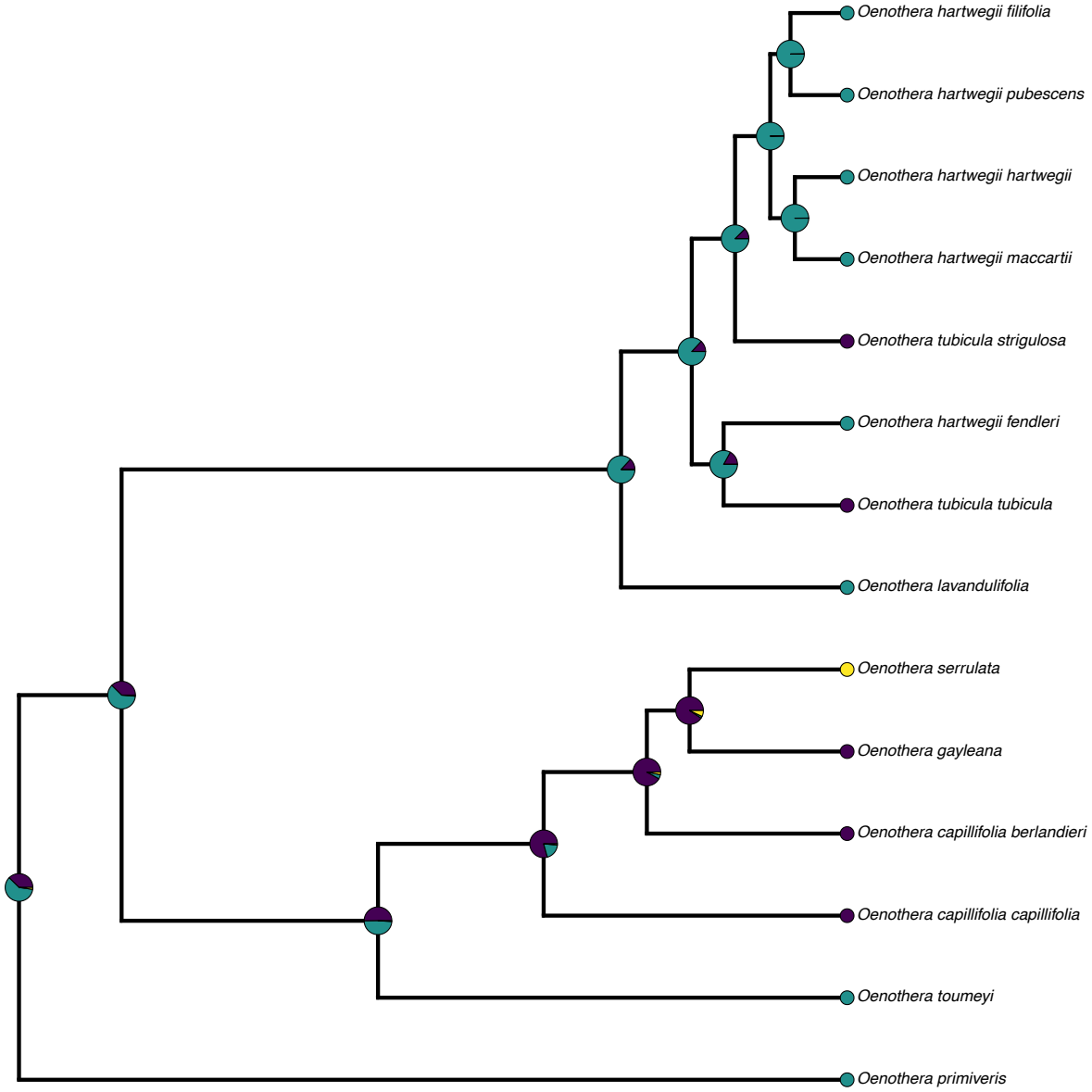
1094

1095 **S11**
1096 See excel table "S11 Morphometric Data"
1097

1098 **S12**
1099 See excel table "S12 Pollen Counts"
1100

1101 **S13**
1102 DNA EXTRACTION PROTOCOL – see additional file

1103 **S14**



1104
1105 Ancestral State Reconstruction of reproductive system in sect. *Calylophus* using supercontigs
1106 and accessions grouped into taxa. Pie-charts on notes represent likelihood of ancestral
1107 reproductive system at each node (teal = hawkmoth pollination, purple = bee pollination, yellow
1108 = PTH).